A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets

RMetS
Royal Meteorological Society

A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets

A. J. Simmons,* P. Berrisford, D. P. Dee, H. Hersbach, S. Hirahara† and J.-N. Thépaut
European Centre for Medium-Range Weather Forecasts, Reading, UK

*Correspondence to: A. J. Simmons, ECMWF, Shinfield Park, Reading RG2 9AX, UK. E-mail: adrian.simmons@ecmwf.int

The ERA-Interim and JRA-55 reanalyses of synoptic data and several conventional analyses of monthly climatological data provide similar estimates of global-mean surface warming since 1979. They broadly agree on the character of interannual variability and the extremity of the 2015/2016 warm spell to which a strong El Niño and low Arctic sea-ice cover contribute. Nevertheless global and regional averages differ on various time-scales due to differences in data coverage and sea-surface temperature analyses; averages from those conventional datasets that infill where they lack direct observations agree better with the averages from the reanalyses. The latest warm event is less extreme when viewed in terms of atmospheric energy, which gives more weight to variability in the Tropics, where the thermal signal has greater vertical penetration and latent energy is a larger factor.

Surface warming from 1998 to 2012 is larger than indicated by earlier versions of the conventional datasets used to characterize what the Fifth Assessment Report of the Intergovernmental Panel on Climate Change termed a hiatus in global warming. None of the datasets exhibit net warming over the Antarctic since 1979.

Centennial trends from the conventional datasets, HadCRUT4 on the one hand and GISTEMP and NOAAGlobalTemp on the other, differ mainly because sea-surface temperatures differ. Infilling of values where direct observations are lacking is more questionable for the data-sparse earlier decades. Change since the eighteenth century is inevitably more uncertain than change over and after a modern baseline period. The latter is arguably best estimated separately for taking stock of actions to limit climate change, exploiting reanalyses and using satellite data to refine the conventional approach. Nevertheless, early in 2016 the global temperature appears to have first touched or briefly breached a level 1.5 °C above that early in the Industrial Revolution, having touched the 1.0 °C level in 1998 during a previous El Niño.

Key Words: reanalysis; temperature trends; Arctic warming; El Niño; atmospheric energy

Received 6 June 2016; Revised 30 September 2016; Accepted 17 October 2016; Published online in Wiley Online Library 5 December 2016

1. Introduction

The latest two assessment reports of the Intergovernmental Panel on Climate Change (IPCC) have stated that warming of the climate system is unequivocal, citing among other evidence the increases in global average surface air and ocean temperatures inferred from observations (IPCC, 2007, 2013). Nevertheless differences in estimates of short-term trends in global-mean surface temperature have been sufficiently large to prompt debate within the scientific community over reference to a 'hiatus' or 'slowdown' in warming over the 15 or so years following the 1997/1998 El Niño event (Lewandowsky *et al.*, 2015; Fyfe *et al.*, 2016). Differences between datasets in their rankings of individual years and months in terms of warmth also hamper clear public communication of reliable information concerning extremes.

These issues have come to the fore over the past year for several reasons. As discussed in the body of this article, newer versions and a wider range of datasets show a higher rate of warming from 1998 to 2012 than in the datasets available to and analysed by IPCC (2013). The global average surface air temperature also reached a level early in 2016 which is by a considerable margin unprecedented over the period of instrumental record.

†Contribution made while on secondment from the Japan Meteorological Agency.

Moreover, the aim of 'holding the increase in the global average temperature to well below 2 °C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5 °C' was agreed by nations meeting in Paris late in 2015 (UNFCCC, 2015). The Paris Agreement includes an undertaking to take stock periodically of progress towards achieving its purpose and long-term goals, starting in 2023 and continuing at intervals of 5 years unless subsequently decided otherwise. This points to a continuing need to reduce uncertainties in estimates of surface temperature changes and to improve the interpretation of the sub-decadal variations in a temperature record that combines interacting effects of anthropogenic and natural external forcings and internal variability of the climate system.

Aside from setting its target for limiting the rise in global temperature, the Paris Agreement also established a global goal of enhancing capacity to adapt to climate change. In doing so, it recognised that the challenge of adaptation had local, subnational, national, regional and wider international dimensions, and required strengthened research, systematic observation and early warning systems. This in turn sets requirements for the observation, analysis and prediction of key impact variables, including needs for spatial and temporal resolution and estimates of uncertainty. Comprehensive reanalyses that use fixed modern data assimilation systems to synthesize states of the atmosphere and interacting components of the climate system from past and present observations have the potential to make a substantial contribution to satisfying these requirements. They provide globally complete estimates of many of the key variables with a frequency and resolution that are becoming increasingly high in newer products, but also provide global and regional averages that can be competitive with those from conventional monthly temperature and humidity products for identifying trends and low-frequency variability over recent decades.

This article presents a new assessment of large-scale temperature trends and variability from reanalyses and conventional monthly surface climatological datasets. The latter include the latest available versions of the three datasets most widely used in IPCC assessments and in the annual statements of the World Meteorological Organization (WMO) on the status of the global climate. The article also documents the representations by the various datasets of the extreme positive temperature anomalies that have occurred over the past 12 or more months. It explores the extent to which values created by infilling or extrapolation in what are observation-void areas for the conventional datasets agree with the values produced in such regions by reanalyses, which infer values there from additional synoptic observations of surface air temperature, from *in situ* and satellite observations of other variables, and from modelling. Sources of differences among datasets are identified. A particular aim is to establish more firmly the credentials of reanalysis for monitoring global and regional temperature, especially with regard to use of European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis (ERA) products to provide information delivered by Europe's Copernicus Climate Change Service (http://climate.copernicus.eu). (All web sites mentioned were accessed on 4 November 2016.) This necessitates placing recent decades in the centennial-scale context for which the conventional datasets still play the primary role. Results from reanalysis are also used to discuss variations in atmospheric energy and mid- to upper-tropospheric temperature, both of which also reached record levels in 2016, but by a smaller margin than in the case of surface temperature.

The outline of the article is as follows. Information on the datasets and their processing is given in the following section. Time series of anomalies in global- and European-average temperatures from 1979 onwards are compared in section 3, and global trends are discussed in section 4. Differences in individual monthly values are examined in section 5. Maps illustrating geographical coverage and variations are presented in section 6, and the contributions of the polar regions and middle and low latitudes to global trends and variability form the topic of section 7. The period from 1979 is placed in longer-term context in section 8, while section 9 discusses variations in atmospheric energy and upper-air temperature. Concluding discussion is provided in section 10.

2. Datasets

2.1. *The choice of datasets*

This article updates and extends several earlier evaluations of ERA surface air temperature products. Those evaluations included comparisons with one or more other reanalyses, conventional climatological datasets and direct observations (Simmons *et al.*, 2004, 2010; Jones *et al.*, 2012; Simmons and Poli, 2015). For this study it was decided as before to compare only a limited number of the available global datasets. This allowed a more thorough investigation and presentation of results from the selected datasets, avoided a mixing of older and newer datasets that could have clouded the conclusions, and enabled some of the interdependencies between datasets to be filtered out or studied in a controlled way.

Most notable among the dataset interdependencies is using the same analysis of sea-surface temperature (SST). Others that are readily identifiable include using common observational input streams or common methodologies. Nevertheless it is difficult without performing comparisons to know quite how independent two datasets are. Two reanalyses may use a largely similar set of observations and the same basic analysis method, but they are likely to differ in the observational quality control and error specifications used in their data assimilation, and in the error characteristics of their assimilating models. Differences in boundary-layer and surface parametrizations, for example, may cause significant differences in surface air temperature over regions where there are insufficient direct observations to constrain the analyses.

Accordingly a selection was made of the latest available versions of widely used products that were known or expected to be of good quality and that had one or more features that distinguished them from other datasets. A central set of four datasets was chosen, two reanalyses and two conventional datasets, on the basis that they incorporate different SST analyses. The conventional datasets were complemented by choosing two other datasets that differed in the completeness of their global coverage, despite having much in common with one or other of the original pair. A third reanalysis was also examined. The datasets and some of their interdependencies are discussed in the following two sub-sections.

The latest versions of datasets as available on 19 September 2016 are used. They comprise values for months up to and including July 2016 in all cases, and include values for August 2016 in the case of the reanalyses and one of the conventional datasets.

2.2. *Reanalyses*

The two reanalyses for which comprehensive results are presented are ECMWF's ERA-Interim (Dee *et al.*, 2011; see Appendix for data access) and the Japan Meteorological Agency's JRA-55 (Kobayashi *et al.*, 2015; http://jra.kishou.go.jp/JRA-55/index_en.html#download), which cover the time ranges from January 1979 and January 1958 respectively. JRA-55 uses the COBE SST reanalysis (Ishii *et al.*, 2005), which is based only on *in situ* measurements, whereas the sequence of SST analyses used by ERA-Interim benefits from observations from satellites as well as direct measurements. Both reanalyses assimilate many types of observation in their 4D-Var schemes, and the data presented for assimilation are largely the same for each. The synoptic screen-level temperature data analysed by both come from many more stations than provide the monthly data used in the conventional analyses, but a few of the stations that report monthly do not provide synoptic data that regularly reach ECMWF via the WMO

Global Telecommunications System. This includes some stations in data-sparse parts of Africa. Maps and statistics can be found in GCOS (2015).

Monthly averages are used for comparison with the conventional datasets. Both ERA-Interim and JRA-55 provide a 2 m temperature product from optimal-interpolation analyses of screen-level observations, using background fields provided by their main 4D-Var data assimilation schemes. Except where stated otherwise, this study uses the analysis fields over land and the background fields over sea. For ERA-Interim this is because its analyses have a warm daytime bias over sea due to use of ship observations that are not adjusted to account for the unrepresentative nature of the daytime measurements due to solar heating, as was the case for the earlier ERA-40 reanalysis (Simmons *et al.*, 2004). As there are changes over time in both ship size and the amount of ship data, trends from ERA-Interim's marine temperature analyses cannot be regarded as reliable. Using background rather than analysis values over sea also gives more consistent global trends in the case of JRA-55.

A further adjustment of the standard ERA-Interim output is made. The 2 m temperatures over ice-free sea, and the SSTs when used alternatively, are reduced by 0.1 °C for all months prior to January 2002. This is to ensure consistency with subsequent temperatures, which are quite uniformly around 0.1 °C cooler due to a change in external source of SST analysis, to which 2 m temperature over sea is closely linked. Many of the results presented here concern the period after 2002, and are in essence unaffected by the offset applied for earlier years. Removing the offset reduces the global ERA-Interim trends quoted in section 4 by 0.03 °C per decade for the full period and 0.05 °C per decade for 1998–2012. Uncertainty in the offset is low enough to introduce uncertainties in ERA-Interim trends that are close to or below the two-figure precision of the quoted values, and appreciably smaller than the differences in trends between the various datasets. Further information on the offset can be found in Simmons and Poli (2015) and cited references, and in the Appendix to this article.

Previous comparisons of ERA-Interim and JRA-55 for upper-air temperature (Simmons *et al.*, 2014) and Arctic surface air temperature (Simmons and Poli, 2015) included evaluations of the MERRA reanalysis (Rienecker *et al.*, 2011). Production of MERRA has since been discontinued, following availability of MERRA-2 (Bosilovich *et al.*, 2015), which runs from 1980 onwards. Monthly-mean 2 m temperature fields* from MERRA-2 have been processed and compared with the other datasets evaluated here. However, results are not discussed at length, as several issues render this reanalysis a clear outlier in terms of trends. This is illustrated later, in sections 7.1 and 7.3.

Other technical aspects of the processing of the reanalysis data are as described by Simmons *et al.* (2010, 2014).

2.3. *Monthly surface climatological datasets*

The central two conventional datasets based on differing SSTs are HadCRUT4 (Morice *et al.*, 2012), produced by the Met Office Hadley Centre in collaboration with the Climatic Research Unit of the University of East Anglia, and NOAAGlobalTemp (Karl *et al.*, 2015), produced by the US National Oceanic and Atmospheric Administration (NOAA). These datasets combine analyses of climatological reports of monthly-mean surface air temperature from stations over land with the producers' own monthly analyses of SST: HadSST3 (Kennedy *et al.*, 2011) for HadCRUT4 and ERSSTv4 (Huang *et al.*, 2015) for NOAAGlobalTemp. Both these SST analyses are based only on *in situ* measurements. SST is used rather than marine surface air temperature as the latter is more difficult to analyse reliably directly from observations. HadCRUT4 and NOAAGlobalTemp analyse monthly data records from many common land stations, but differ in their data collection and quality control (Jones *et al.*, 2012; Gleason *et al.*, 2015).

HadCRUT4 covers the period from January 1850. Data for the latest month are added regularly, but values for previous months are updated only intermittently. Version 4.5.0.0, downloaded from http://www.metoffice.gov.uk/hadobs/), is used in this study. NOAAGlobalTemp runs from January 1880 onwards. Past values in this dataset may change with each monthly release: version 4.0.1.201607, downloaded from https://www.ncdc.noaa.gov/data-access), is used here. Both datasets provide values for $5^\circ \times 5^\circ$ grid squares, and for both there are gaps in global coverage. As illustrated later, these gaps are larger in the case of HadCRUT4, for which no infilling of data is performed to construct values for grid squares for which a direct calculation cannot be made. A gap can occur either because land-station or marine temperature data are lacking for the month in question or because there are insufficient data to establish a background climatological value for the location.

HadCRUT4 is an ensemble of 100 possible realizations of past temperature change that sample some of the uncertainties in estimating multi-decadal variability. Results presented here are primarily for the medians of the values from the ensemble, although the spread of the ensemble is also examined.

Comparisons are also made with the GISTEMP (Hansen *et al.*, 2010) dataset produced by the US National Aeronautics and Space Administration (NASA). GISTEMP covers the period since 1880 and its current version uses largely the same NOAA collection of input land-station observations and ERSSTv4 dataset as NOAAGlobalTemp. Analysed data are provided on a $2^\circ \times 2^\circ$ grid, and global coverage for recent decades is close to, but not quite, complete. This coverage is achieved by using an analysis method that exploits the correlation of temperature change found for stations separated by up to 1200 km (Hansen and Lebedeff, 1987) and data are available generally with 1200 km smoothing. $2^\circ \times 2^\circ$ grid values over many continental land areas and islands are also supplied with 250 km smoothing.

Except where stated, the GISTEMP global means examined here are downloaded as given, for consistency with values quoted by the data provider. Regional means and global maps are produced from a merged dataset that uses $2^\circ \times 2^\circ$ values from the 250 km smoothed dataset where available and 1200 km smoothed values otherwise. This provides detail over land that makes for a fairer local comparison with the reanalyses, and distinguishes GISTEMP more substantially from NOAAGlobalTemp than would be the case had only the 1200 km smoothed data been used. The global means derived from the merged dataset do not match exactly the directly downloaded global means, but differences are quite small, as illustrated later. The GISTEMP datasets for past months may change when a new monthly release is made: the datasets used here apply to the release that contains data up to August 2016, and were downloaded from http://data.giss.nasa.gov/gistemp/.

Two datasets that spatially extend the HadCRUT4 median (Cowtan and Way, 2014) are also examined. The first uses solely the spatial interpolation method known as kriging, while the second is a hybrid approach that adds information from the record of tropospheric temperature derived at the University of Alabama in Huntsville (UAH; Christy *et al.*, 2007) from space-based microwave sounding data. Updated versions of the Cowtan and Way datasets, derived from HadCRUT4 version 4.5.0.0 and downloaded from (and documented at) http://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html, are used in the present study: the hybrid dataset Had4_UAH_v2 that is restricted to the period from 1979 when the satellite data are available and the Had4_krig_v2 dataset covering the period from 1850.

*Downloaded from the MERRA-2 instM_2d_asm_Nx data stream, version 5.12.4, doi: 10.5067/5ESKGQTZG7FO held by the NASA Goddard Space Flight Center Distributed Active Archive Center.

2.4. *Adjustment to a common reference period*

The HadCRUT4 and related datasets comprise values that are anomalies relative to the reference period 1961–1990, NOAAGlobalTemp provides anomalies relative to 1971–2000 and GISTEMP anomalies relative to 1951–1980. A common reference period 1981–2010 is used here. Global or regional averages shown as time series are first calculated for the conventional datasets using their original data values, and values are then adjusted separately for each month to be relative to the 1981–2010 average for the month. The changes in long-term trends and variability from month to month resulting from this adjustment are very small. The times series for the reanalyses are calculated directly from anomalies relative to 1981–2010.

The maps shown of monthly and annual anomalies adjusted to be relative to 1981–2010 present values only for grid squares where there are data for at least 90% of the months from 1981 to 2010, and only for grid squares at which a value is available every month to calculate an annual anomaly. Geographical coverage is somewhat poorer than if values from the conventional datasets had been plotted relative to their native reference periods, particularly in the case of the monthly maps from HadCRUT4. The grid squares for which coverage is lost are indicated in the maps. Apart from this, the adjustment of the baseline causes little change to the appearance of the maps, as the shifts are generally by at most a few tenths of a degree, much less than the amplitude of the spatial variability shown on the maps. Computing global and regional averages of values that are first adjusted to be relative to 1981–2010 for each grid square gives time series that are not precisely the same as those presented here for the conventional datasets, but it has been confirmed for HadCRUT4 that the differences in global averages are much smaller than differences from month to month and between datasets for a particular month.

3. Time series of global and European average temperatures

Figure 1 shows time series of 12 month running averages of the estimated global-mean surface temperature from 1979 onwards for ERA-Interim, JRA-55, the HadCRUT4 median and NOAAGlobalTemp. Each dataset provides a similar overall picture: the general warming since the late 1970s is not in doubt, nor is the occurrence of warmer and colder spells linked with El Niño events, volcanic eruptions, variations in sea-ice cover and other sources of variability. All datasets show above-average values from 2001 onwards, and for each of them the warmest 16 calendar years are 1998 and then 2001 to 2015. In each case, the temperature for the calendar year of 2015 is higher than for any earlier 12 month period, but increasingly exceeded by the 12 month averages ending in each of the first seven months of 2016. Values for 2015 are around 0.45 °C warmer than the 1981–2010 average, with slightly lower values from the reanalyses: 0.43 °C from JRA-55 and 0.44 °C from ERA-Interim, compared with 0.47 °C from NOAAGlobalTemp and 0.48 °C from HadCRUT4. The averages for the 12 months to July 2016 are 0.57 °C for NOAAGlobalTemp, 0.58 °C for HadCRUT4, 0.61 °C for JRA-55 and 0.62 °C for ERA-Interim.

The datasets differ more considerably in their estimates of the magnitudes of individual warm and cold spells, and accordingly in their rankings of the set of warmest calendar years. ERA-Interim shows the largest peaks. Its averages for the calendar years of 2005, 2006 and 2010 all exceed its 2014 average. Conversely, JRA-55 has 2014 as the second warmest calendar year by a narrow margin, although it does have a slightly warmer 12 month spell in 2009–2010. NOAAGlobalTemp shows the lowest maxima in the period from 1999 to 2013; 2014 is clearly the warmest calendar year prior to 2015 for this particular dataset. The temperature anomaly for 2005 ranges from 0.23 °C for NOAAGlobalTemp to 0.35 °C for ERA-Interim. The spread among datasets for this year is the largest for any calendar year in the period. The largest spread in the sets of 12 month means is 0.14 °C, for the mean from May 2005 to April 2006. Spreads above 0.1 °C occur only in the early 1980s and the mid-2000s. Factors behind these differences are discussed in subsequent sections.

HadCRUT4, NOAAGlobalTemp and the related GISTEMP dataset comprise a combination of surface air temperature data over land and SST data. Most reanalysis results presented here are for surface air temperature over both land and sea, but Figures 1(a) and (b) show in dark grey the differences between these global means and global means based on surface air temperature over land and sea-ice, and SST otherwise. Differences are very small, but have a systematic component: they are slightly negative over most of the first half of the period and slightly positive over most of the second half, for both reanalyses. They can also shift the ordering of years according to their warmth: using SST rather than air temperature makes 2006 slightly cooler rather than warmer than 2014 for ERA-Interim. The somewhat larger trend in air than sea temperature is such as to reduce air–sea differences over time, and is consistent with climate-model results reported by

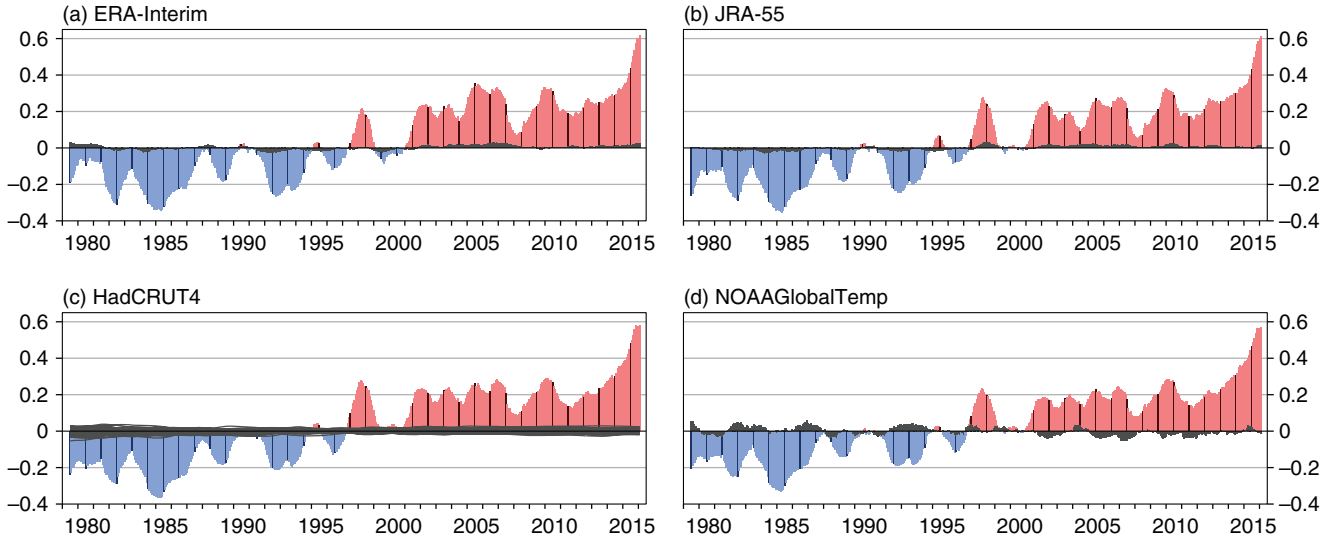

Figure 1. Twelve-month running means of anomalies in globally averaged surface temperature (°C) relative to 1981–2010, from (a) ERA-Interim, (b) JRA-55, (c) the HadCRUT4 median and (d) NOAAGlobalTemp, for January 1979 to July 2016. Pink denotes above-average values and blue below-average values. The darker bars are calendar-year means. ERA-Interim and JRA-55 values are based on surface air (2 m) temperature over sea; the small differences between these values and those using sea-surface temperature (as in HadCRUT4 and NOAAGlobalTemp) are shown in dark grey in (a) and (b). The overlapping dark grey lines in (c) denote the differences between the values of the 100 HadCRUT4 ensemble members and the HadCRUT4 median. The differences between NOAAGlobalTemp and GISTEMP are shown in grey in (d).

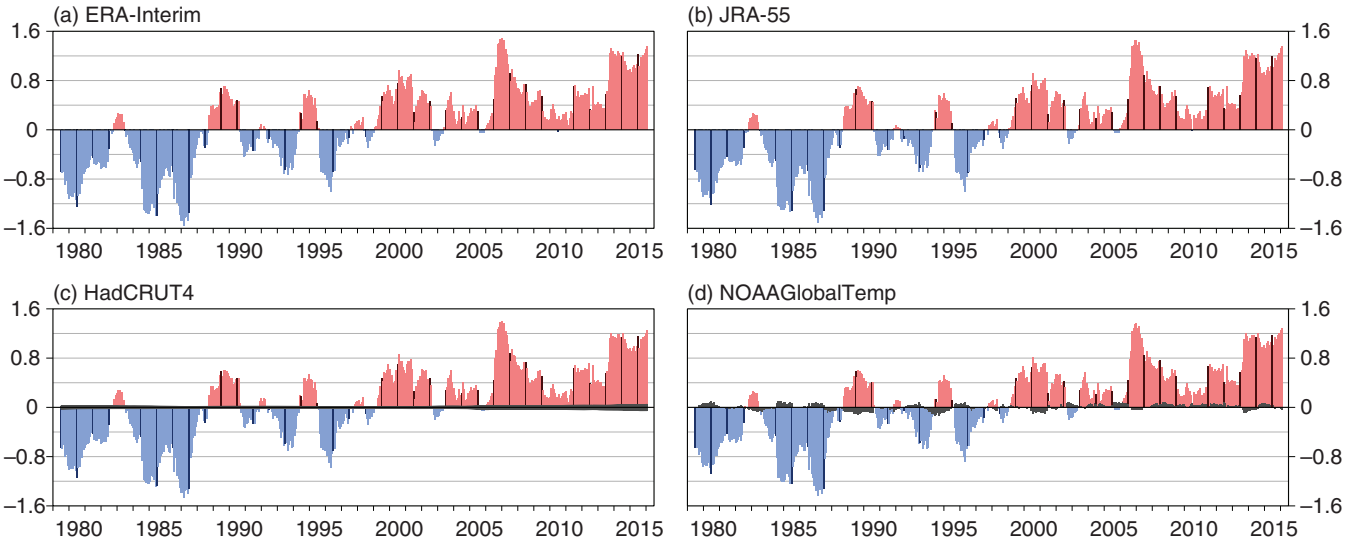

Figure 2. Twelve-month running means of anomalies in surface air temperature over Europe (°C), relative to 1981–2010, from (a) ERA-Interim, (b) JRA-55, (c) the HadCRUT4 median, and (d) NOAAGlobalTemp, for January 1979 to July 2016. Colour shading and darker bars are as in Figure 1. Dark grey lines are plotted in (c) to denote the differences between the values of the 100 HadCRUT4 ensemble members and the HadCRUT4 median; they all lie close to the zero line. Differences between NOAAGlobalTemp and GISTEMP values are shown in grey in (d). Values are averages over land areas located between 20°W and 40°E, and 35°N and 80°N. The ERA-Interim land-sea mask is used to partition coastal grid-box values between land and sea.

Cowtan *et al.* (2015) and Huang *et al.* (2015). Although small, the differences do need to be kept in mind when comparing reanalyses with the conventional surface-temperature datasets, as is the case also when model outputs are compared with such datasets (Cowtan *et al.*, 2015).

The differences between the HadCRUT4 median and the 100 individual ensemble members (each expressed as an anomaly with respect to its own 1981–2010 average) are plotted as a set of largely overlapping grey lines in Figure 1(c). There is little variation over time in the spread of the ensemble for the period shown. The means for 2015 range from 0.46 to 0.51 °C; those for 2005 range from 0.25 to 0.29 °C. This is despite a considerable difference between the 2 years in the level of agreement among datasets. Indeed, the spread of the HadCRUT4 ensemble is generally smaller than the spread of the alternative datasets over the period from January 1979. The root-mean-square spread of monthly values for this period is 0.05 °C between ensemble members and 0.10 °C between GISTEMP, the HadCRUT4 median, NOAAGlobalTemp and the two reanalyses. Morice *et al.* (2012) acknowledge that HadCRUT4 does not provide a full description of uncertainties, and advocate that users of the dataset test the robustness of their results by comparing with other datasets. In particular, all members of the HadCRUT4 ensemble have the same limited data coverage, so by themselves provide no information on uncertainties in global averages that arise from this. Morice *et al.* (2012) provide estimates of additional uncertainty in their calculations of temporal and spatial averages, arising from measurement uncertainty, under-sampling within a grid box and (using reanalysis data) lack of coverage. These estimates are not included in the results presented here.

Figure 1(d) includes the differences between NOAAGlobalTemp and GISTEMP values, shown in grey. Although generally small compared to the variations of either, there is a systematic component to the differences: GISTEMP has generally larger negative anomalies earlier in the period and larger positive anomalies later in the period. 2015 is an exception in that GISTEMP has a smaller anomaly of 0.44 °C, the same as ERA-Interim. GISTEMP is generally closer to the reanalyses than HadCRUT4 and NOAAGlobalTemp are. This suggests that the greater geographical extent of the values it provides is reasonably consistent with what is provided by the reanalyses, as will be seen in specific cases discussed later. The same holds for the extensions of HadCRUT4 provided by Had4_UAH_v2 and Had4_krig_v2, also illustrated later.

Figure 2 shows time series of temperatures averaged over European land areas. Variability is much higher for this considerably smaller domain, but the region is well observed, and all datasets are in good agreement. The reanalyses exhibit very slightly larger maxima and minima than the other datasets, as do HadCRUT4 and GISTEMP compared with NOAAGlobalTemp. For Europe, 2014 and 2015 are the warmest two calendar years on record, with little to separate them. However, neither calendar year is quite as warm as the 12 months from the middle of 2006 to the middle of 2007.

4. Global temperature trends

Linear trends in global-mean surface temperature are presented in Figure 3. They have been computed by least-squares fits to monthly anomalies relative to 1981–2010; these anomalies are also shown. The latest warm spell is exceptional in the extent to which it deviates from the linear trend over the full period from January 1979 to July 2016, for all datasets. The deviation is close to 0.5 °C in February 2016 for ERA-Interim, JRA-55 and GISTEMP.

The full-period trends differ little among the datasets. ERA-Interim, JRA-55 and GISTEMP give warming rates (all rates given as °C per decade) of 0.17, HadCRUT4 gives 0.18 while NOAAGlobalTemp gives 0.16. Trends range from 0.17 to 0.19 for the HadCRUT4 ensemble, and are 0.19 for Had4_UAH_v2 and 0.18 for Had4_krig_v2. The trends from both reanalyses are slightly smaller if SST rather than marine air temperature is used, rounding to 0.17 again for ERA-Interim but to 0.16 for JRA-55. Trends are slightly smaller still if the reanalyses' background forecasts of air temperature are used over land rather than the values from analysing screen-level observations: in this case the trends round to 0.16 for both reanalyses.

The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (AR5; IPCC, 2013) reported a warming rate of 0.05 with a 90% uncertainty range from −0.05 to 0.15 for the 15 year period 1998–2012, compared with a rate of 0.12 with uncertainty range from 0.08 to 0.14 for the period 1951–2012. AR5 refers to this as 'the hiatus in global mean surface warming of the past 15 years', although the use of words such as 'hiatus' or 'slowdown' has caused debate (Lewandowsky *et al.*, 2015; Fyfe *et al.*, 2016). Results from reanalyses were not used in computing these trends, and newer versions or replacements are now available for the three datasets (HadCRUT4, NOAA's MLOST (Vose *et al.*, 2012) and GISTEMP) which were used in the AR5 calculation.

The datasets examined here unsurprisingly differ much more in their trends estimated for 1998–2012 than they do in their

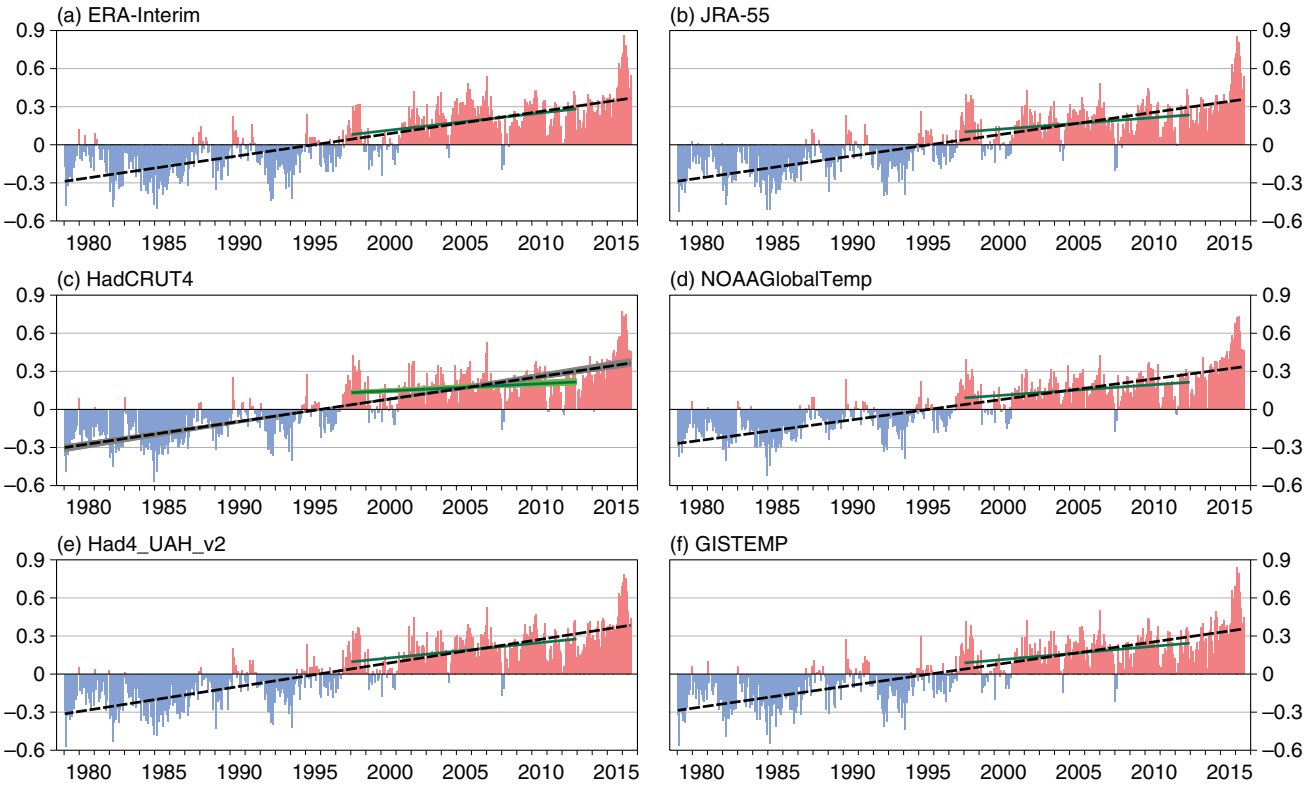

Figure 3. Monthly anomalies in globally averaged surface temperature (°C) relative to 1981–2010, from (a) ERA-Interim, (b) JRA-55, (c) the HadCRUT4 median, (d) NOAAGlobalTemp, (e) Had4_UAH_v2 and (f) GISTEMP, for January 1979 to July 2016. Also shown are least-squares linear fits to the monthly values computed for the full period (black, dashed lines) and for 1998–2012 (dark green, solid lines). In the case of HadCRUT4, the corresponding linear fits for each ensemble member are plotted as sets of (overlapping) grey and lighter green lines.

trends for the full period. Apart from 12 HadCRUT4 ensemble members, all datasets give 1998–2012 trends that are higher than AR5's central estimate, although all lie within its uncertainty interval. All are lower than the trend values for the full period, but most are within the range of the 1951–2012 trend quoted in AR5. The 1998–2012 warming rate from the HadCRUT4 median is 0.06, close to the central estimate from the IPCC report, while the HadCRUT4 ensemble members range from 0.04 to 0.08. MLOST gave a rate of 0.04, but its NOAAGlobalTemp replacement provides a higher value of 0.08. Karl *et al.* (2015) discuss the reasons for this higher estimate. The 15 year warming rates from the other datasets are higher still: 0.09 for JRA-55, 0.10 for GISTEMP, 0.11 for Had4_krig_v2, 0.12 for Had4_UAH_v2 and 0.14 for ERA-Interim. The ERA-Interim trend may be overestimated due to its use of an SST analysis that underestimated the warmth of the 1997/1998 El Niño (Simmons and Poli, 2015), but the overestimate from this cause appears to be by no more than 0.01. This is based on repeating the calculations using SST rather than marine air temperature, employing the combination of HadISST2 (version 2.1.1.0; J. J. Kennedy *et al.*, 2016; personal communication) and OSTIA (Donlon *et al.*, 2012) SST analyses proposed by Hirahara *et al.* (2016) for use in ERA5, ECMWF's latest comprehensive atmospheric reanalysis which is currently in production as the replacement for ERA-Interim.

5. Differences in monthly values

Variations from month to month in global averages can be seen in Figure 3. Most notable is the extremity of the positive temperature anomalies in late 2015 and early 2016. Values from the reanalyses and GISTEMP reach around 0.85 °C above the 1981–2010 norm in February 2016.

All datasets identify January 2007 as the warmest month (relative to its 1981–2010 norm) preceding the latest warm spell. The anomaly for this month ranges from 0.43 °C for NOAAGlobalTemp to 0.54 °C for ERA-Interim. In all datasets, the January 2007 maximum is exceeded by the monthly temperature anomalies for each month from October 2015 to April 2016.

Nevertheless there are quite pronounced variations between some of the datasets for particular months within the recent extreme spell. Table 1 documents values from October 2015 to July 2016 for all datasets. For several months in this spell the

Table 1. Anomalies in global-mean surface temperature (°C) relative to 1981–2010 for the months of October 2015 to July 2016 from the datasets listed in the left column.

Dataset	October 2015	November 2015	December 2015	January 2016	February 2016	March 2016	April 2016	May 2016	June 2016	July 2016
ERA-Interim	0.64	0.60	0.69	0.72	0.86	0.78	0.69	0.59	0.44	0.55
JRA-55	0.63	0.57	0.68	0.72	0.85	0.81	0.70	0.56	0.44	0.54
HadCRUT4 median	0.57	0.58	0.77	0.58	0.73	0.75	0.62	0.42	0.47	0.46
HadCRUT4 ensemble	0.55–0.59	0.56–0.61	0.75–0.80	0.56–0.61	0.72–0.76	0.73–0.78	0.60–0.65	0.41–0.45	0.44–0.49	0.44–0.49
Had4_UAH_v2	0.63	0.61	0.69	0.72	0.78	0.76	0.65	0.50	0.39	0.44
Had4_krig_v2	0.65	0.63	0.70	0.72	0.79	0.76	0.63	0.48	0.40	0.42
NOAAGlobalTemp	0.58	0.55	0.68	0.60	0.72	0.73	0.62	0.45	0.48	0.46
GISTEMP	0.66	0.60	0.69	0.69	0.84	0.79	0.64	0.52	0.40	0.44
	(0.65)	(0.63)	(0.69)	(0.71)	(0.83)	(0.79)	(0.66)	(0.55)	(0.40)	(0.44)

Values for GISTEMP are shown both as derived from downloaded global means and (in brackets) as derived from the merged 2° × 2° dataset discussed in section 2.

HadCRUT4 dataset is a clear outlier, in that its median and all, or almost all, of its ensemble values are either below or above all values from the other datasets. In particular, it gives the least warm values in January and February 2016, but the warmest values in December 2015. There is little doubt that this outlying behaviour is due to the dataset's limited spatial sampling, as the two spatially extended versions of HadCRUT4 are much closer to the other datasets, particularly the two reanalyses, which in turn are close to each other. It is also noteworthy that the more spatially complete GISTEMP dataset, either in original form or as a merge of 250 and 1200 km smoothed values, is for the most part closer to the reanalyses than is NOAAGlobalTemp.

The anomalies for August 2016 from ERA-Interim, JRA-55 and GISTEMP are higher than those for June and July, at 0.62 °C for ERA-Interim, 0.56 °C for JRA-55 and 0.58 °C for GISTEMP. Twelve-month running means from September 2015 are the highest on record: 0.63, 0.62 and 0.60 °C respectively.

6. Geographical coverage

Maps of the annual-mean temperature anomalies relative to 1981–2010 from six datasets are presented in Figure 4 for 2011 and 2015. The year 2011 was chosen in addition to the latest complete calendar year for several reasons. It was a year with an SST anomaly over the Pacific opposite to that in 2015, as can be clearly seen for all datasets in Figure 4. It was also the year with the highest positive temperature anomaly averaged north of 60°N in the datasets with complete coverage, and the year with the largest spread in estimates of global means since the 2005–2007 period which is discussed further in the following section.

The Arctic void in data from HadCRUT4 and NOAAGlobalTemp covers only a small part of the globe, but it is the region where the largest temperature anomalies occur in the reanalyses, associated with anomalous winter sea-ice conditions (Simmons and Poli, 2015). Arctic temperatures much above normal occurred early and late in 2011. Summer temperatures were also relatively warm, though not as warm as in 2007, 2012 and 2016, the only years with lower minimum Arctic sea-ice extent according to the Sea Ice Index of the US National Snow and Ice Data Center (NSIDC; http://nsidc.org/arcticseaicenews). The Arctic warmth of 2011 as a whole is shown clearly by the maps for the reanalyses in Figure 4. High temperatures are also indicated by HadCRUT4 where values are available. NOAAGlobalTemp provides fewer data values over northwestern Russia: for the island of Novaya Zemlya and around the Kara Sea to the east. ERA-Interim has been shown by Simmons and Poli (2015) to fit well the wintertime synoptic data from stations in this region. Nearby values from NOAAGlobalTemp are less anomalous than the values from the reanalyses and HadCRUT4. These differences carry over into the corresponding spatially extended datasets: GISTEMP has lower positive temperature anomalies than Had4_UAH_v2 over the Arctic. The same is seen for 2015.

Elsewhere the patterns and amplitudes of the temperature anomalies shown in Figure 4 are in generally good agreement where observational coverage is good. ERA-Interim and JRA-55 differ most over western and southern Africa, over South America (where it is ERA-Interim that is the more consistent with the conventional datasets) and over Antarctica. HadCRUT4 and NOAAGlobalTemp do not provide values over sea-ice off the coast of Antarctica, and the spatially more extensive

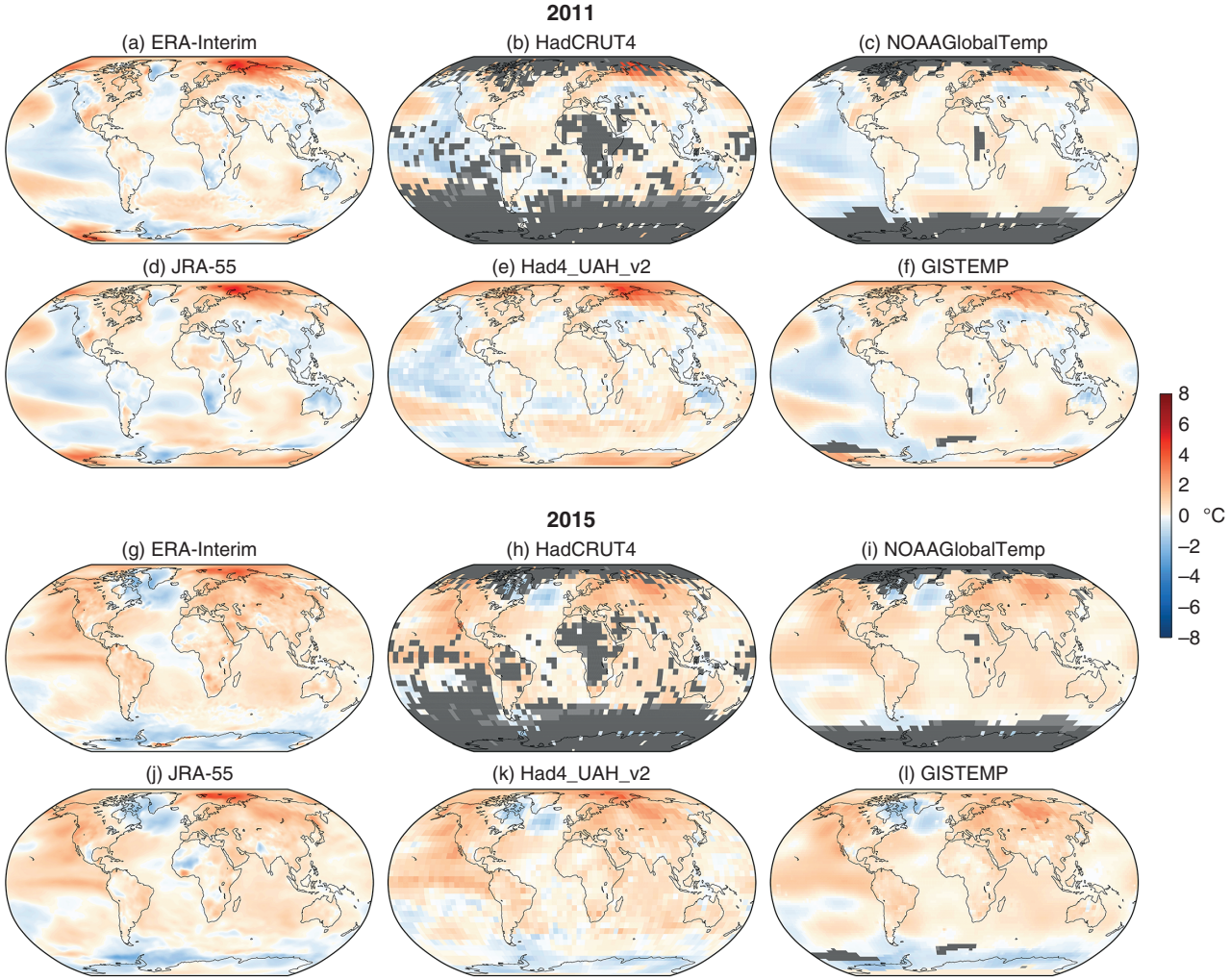

Figure 4. Surface temperature anomalies (°C) relative to 1981–2010 for (a)–(f) 2011 and (g)–(l) 2015, from (a, g) ERA-Interim, (b, h) HadCRUT4, (c, i) NOAAGlobalTemp, (d, j) JRA-55, (e, k) Had4_UAH_v2 and (f, l) GISTEMP. Grid boxes where values are missing are coloured grey. Lighter grey colouring indicates boxes that would have had values had maps been presented as anomalies relative to the standard reference period of each dataset.

Had4_UAH_v2 and GISTEMP datasets have weaker anomalies in this region than the reanalyses, which are reasonably consistent in their depictions of temperature anomalies that can be linked to anomalies in sea-ice cover.

The datasets also differ in their resolution of SST anomalies, most evidently that over the tropical Pacific Ocean associated with the El Niño in 2015. The two reanalyses provide a much sharper picture than NOAAGlobalTemp and GISTEMP, which as noted earlier use the same SST analysis. Although supplied on the same $5° \times 5°$ grid as NOAAGlobalTemp, HadCRUT4 (where it provides data) and the extended Had4_UAH_v2 provide a more detailed depiction of SST anomalies, closer to that of the reanalyses.

Differences from monthly climatological averages for 1981–2010 are illustrated in Figure 5 for December 2015 and January and February 2016. The datasets are in overall agreement for the 3 months, showing relatively warm conditions over the tropical and subtropical oceans, though with declining amplitude over the Pacific as time progresses, and persistent warm conditions over most of South America and southern Africa. Relatively warm conditions also persisted over the Barents and Kara Seas and over the adjacent Arctic Ocean to the north, where winter sea-ice cover was unusually low, as indicated either by the datasets used by ERA-Interim and JRA-55 or by the NSIDC Index. Elsewhere, although the winter was predominantly less cold than average at middle and high northern latitudes, exceptionally so for some regions and months, there was also pronounced month-to-month variability. Temperatures shifted from above to below and then again to above average over eastern Europe for example, and below-average temperatures over parts of the Arctic in December gave way to above-average temperatures in January.

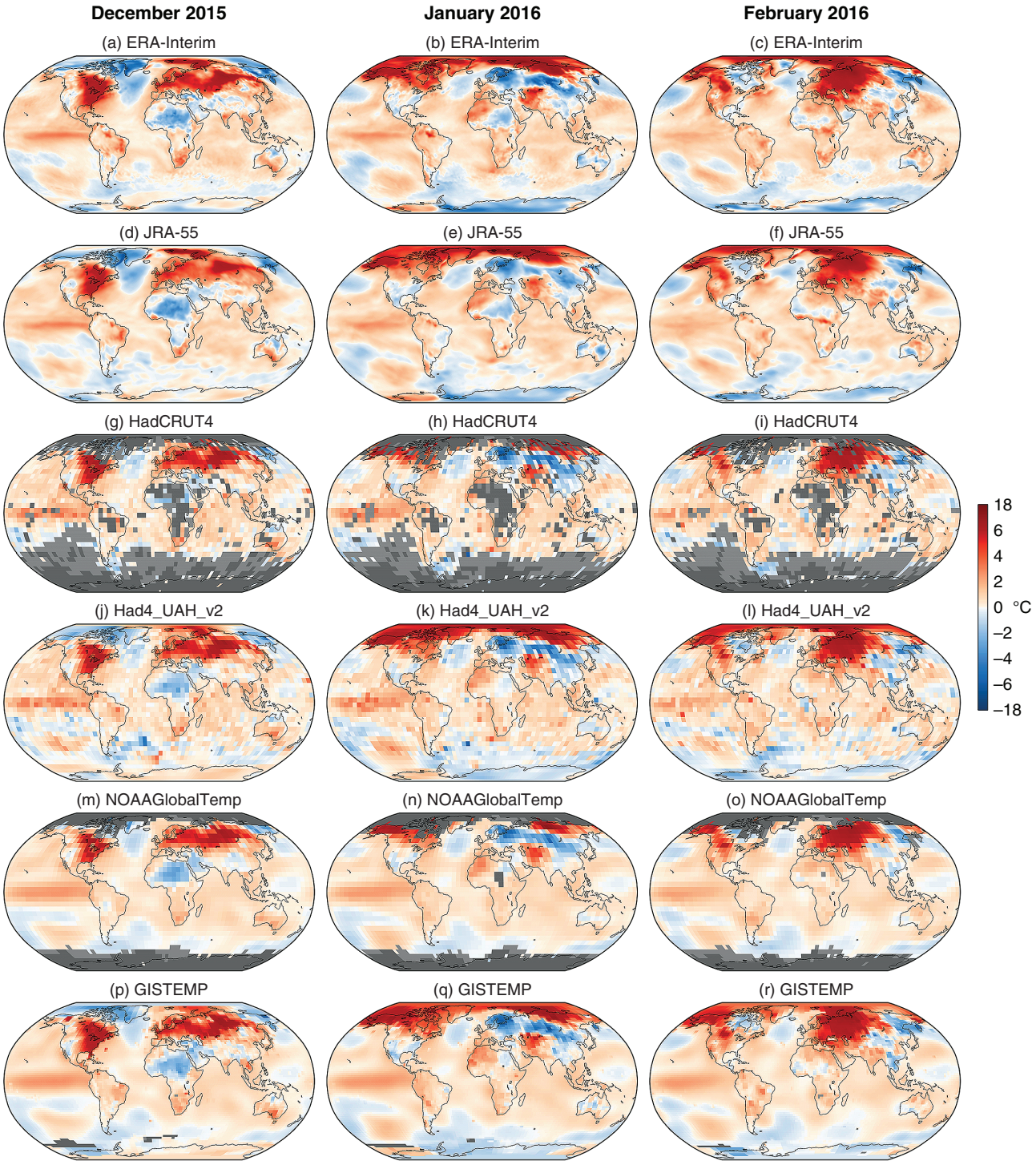

Figure 5. Surface temperature anomalies (°C) relative to 1981–2010 for (left) December 2015, (centre) January 2016 and (right) February 2016, from (a, b, c) ERA-Interim, (d, e, f) JRA-55, (g, h, i) HadCRUT4, (j, k, l) Had4_UAH_v2, (m, n, o) NOAAGlobalTemp and (p, q, r) GISTEMP. Grey shading is as in Figure 4.

The comments made on the resolution provided by the various datasets with regard to the annual-mean maps apply also to the monthly maps. The HadSST3 analysis used in HadCRUT4 can be seen to be prone to produce local values for grid squares that stand out from neighbouring values, and these are inherited by Had4_UAH_v2. These generally do not have counterparts in the fields from the reanalyses. It is beyond the scope of this study to evaluate comprehensively the local detail provided by the reanalyses over land, which is not always in good agreement between the two, but the spatial and temporal variations that are sharper in the reanalyses than in the other datasets over Australia compare reasonably with the anomalies in mean monthly temperatures reported routinely by the Bureau of Meteorology at http://www.bom.gov.au/climate/current/

The principal differences among datasets in the global-mean temperatures for these months shown in Table 1 can be appreciated qualitatively from the differences in spatial coverage of the datasets shown in Figure 5. HadCRUT4's higher mean values in December are consistent with it missing negative anomalies over northern Africa and a quite substantial part of the Arctic. Lower values from both HadCRUT4 and NOAAGlobalTemp in January and February are consistent with them missing above-average temperatures over much of the Arctic. The spatial extensions provided by GISTEMP and Had4_UAH_v2 appear to work well for these months as judged by comparison with the values provided through the radically different approach of reanalysis.

Although the spatially extended datasets generally give better agreement with the reanalyses, this is not invariably the case. Table 1 shows this for July 2016. Maps of monthly anomalies (not included here, but viewable on the websites of some of the dataset providers) are revealing as to the reasons. The reanalyses show only small anomalies over Arctic summer sea-ice and surrounding seas, where surface air temperatures are in agreement with sparse observations that differ little from 0 °C (Simmons and Poli, 2015). In contrast, the spatially extended datasets may spread more-anomalous values from high-Arctic land stations out over sea-ice. Substantial differences are also found for the winter Antarctic, where temperature anomalies in regions of anomalous sea-ice cover that are opposite in sign to the anomalies over the neighbouring Antarctic Plateau are more prevalent in the reanalyses than in the conventional datasets.

7. Contributions to global means from polar and other regions

7.1. *Averages over the polar regions*

A more quantitative identification of differences is provided by time series of temperature anomalies for various regions. Figure 6 shows 12 month running means for the polar regions north of 60°N and south of 60°S where most of the sea-ice cover occurs. These regions are referred to here simply as the Arctic and Antarctic; results differ little if the boundaries are placed at 65° latitude. The full averages over these domains for ERA-Interim and JRA-55 in Figure 6(a) show the substantial Arctic warming that was examined for these reanalyses by Simmons and Poli (2015) for the years up to 2013. The averages for HadCRUT4 and NOAAGlobalTemp based on their partial coverage of the Arctic are lower for much of the period from around 2005 onwards. Corresponding values for the Antarctic show little long-term change.

Figures 6(c) and (d) compare instead the full-domain averages of ERA-Interim and JRA-55 with GISTEMP and Had4_UAH_v2. The better coverage of GISTEMP and Had4_UAH_v2 compared with NOAAGlobalTemp and HadCRUT4 brings better agreement with the reanalyses for the Arctic, especially later in the period and more so for Had_UAH_v2 than GISTEMP. GISTEMP and Had4_UAH_v2 also improve agreement for the Antarctic for much of the period, though not for the latest years, when the two are close to each other, but less so to the reanalyses.

Figures 6(e) and (f) provide corresponding results for CRUTEM4 (Jones *et al.*, 2012), the land component of HadCRUT4, and for the reanalyses when sampled only at the Arctic and Antarctic grid squares where CRUTEM4 provides values. CRUTEM4 is chosen because it provides more data values over land south of 60°S than NOAAGlobalTemp, and values only for

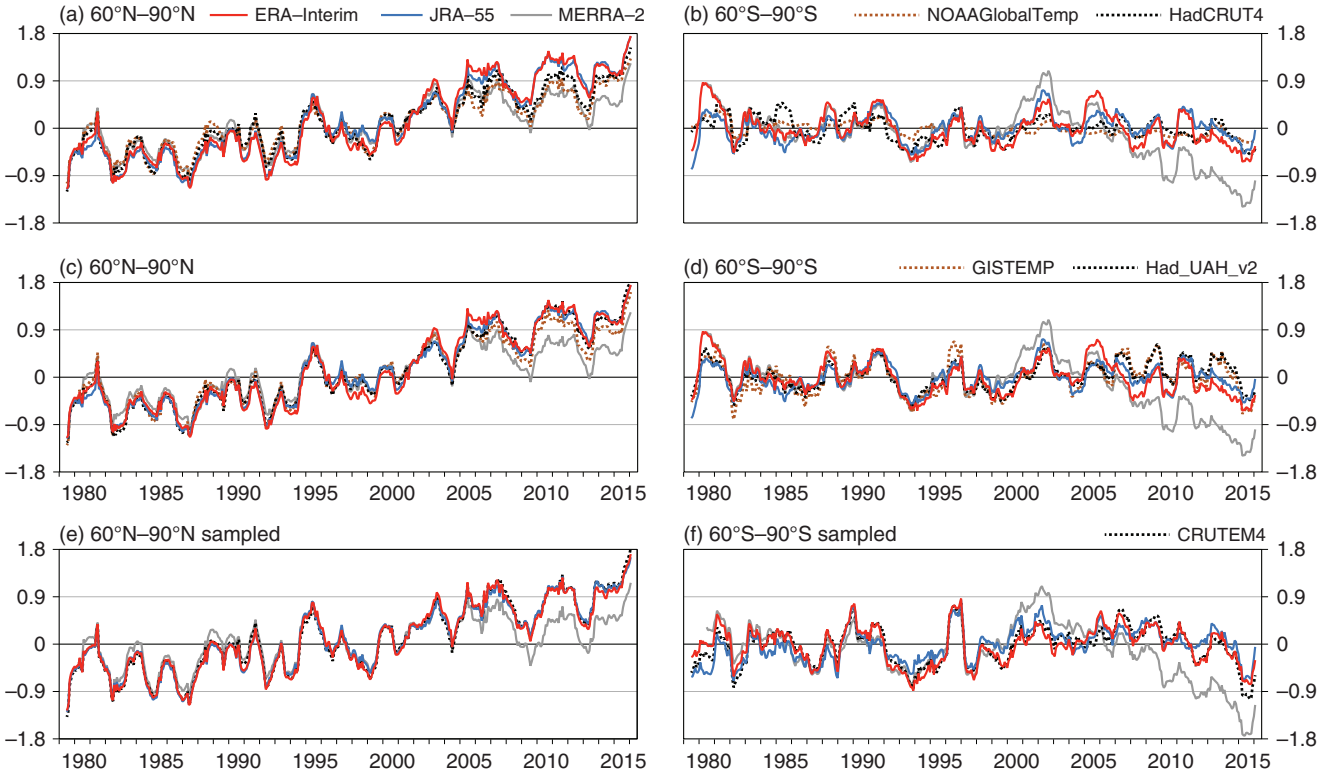

Figure 6. Twelve-month running-mean surface temperature anomalies relative to 1981–2010 (°C) based on data for January 1979 to July 2016 from (a) HadCRUT4 (black, dotted) and NOAAGlobalTemp (orange, dotted) averaged over all grid boxes from 60°N to 90°N where they provide values, and from ERA-Interim (red, solid), JRA-55 (blue, solid) and MERRA-2 (grey solid) averaged over the whole 60°N–90°N domain. (b) is as (a) but for 60°S–90°S, (c) and (d) are as (a) and (b), but showing Had4_UAH_v2 (black, dotted) and GISTEMP (orange, dotted). (e) shows CRUTEM4 (black, dotted; version 4.4.0.0), ERA-Interim, JRA-55 and MERRA-2 averaged over all grid-boxes from 60°N to 90°N where CRUTEM4 provides values. (f) is as (e) but for 60°S–90°S.

grid squares that include monthly station data. ERA-Interim and JRA-55 are both in quite good agreement with CRUTEM4 when sampled in this way, more so for the less data-sparse Arctic. ERA-Interim is closer to CRUTEM4 than JRA-55 is for the Antarctic.

Figure 6 includes results from MERRA-2. Although the most recent of the reanalyses studied here, it is an evident outlier, providing values that are colder relative to 1981–2010 averages than are provided by all other datasets from around 2005 onwards, for both the Arctic and the Antarctic. Similar behaviour for the Arctic was shown by Simmons and Poli (2015) for its predecessor, MERRA. Maps show that the Antarctic cooling in the later years of MERRA-2 is associated with a shift to colder values around the coastline of Antarctica and over the offshore region that is ice-covered in winter. MERRA-2 does not benefit from an analysis of synoptic surface air temperature observations such as is used in ERA-Interim and JRA-55, but this screen-level analysis adds relatively little for these two reanalyses, as their background fields are close to the analysis fields. This is illustrated by Simmons and Poli (2015) for the Arctic in the case of ERA-Interim.

7.2. *Comparison of ERA-Interim with Antarctic station values*

Figure 7 presents direct comparisons of the annual-mean anomalies in surface air temperature relative to 1981–2010 from the ERA-Interim background with observed values from the six Antarctic stations (specifically, stations with WMO identifiers greater than 89 000) for which ERA-Interim has access to data for every month from 1979 to 2015. The information comes from ERA-Interim's 4D-Var data assimilation system. Although the station data are used only in the separate screen-level analysis, they are passed passively through the 4D-Var system, and a record is kept of how well they are fitted by the background forecast. The 4D-Var operates over a 12 h period, and observations are compared with background values to within 15 min of their reported time, although only observations for the standard synoptic hours of 0000, 0600, 1200 and 1800 UTC are used to produce the averages shown in Figure 7. This reduces a possible effect of changes in frequency of reporting, which in general has increased over time (GCOS, 2015). The mean error of the background ranges from 3.6 to −6.2 °C for these stations, all of which are close to the Antarctic coast. Fréville *et al.* (2014) discuss a positive bias in ERA-Interim land- and air-surface temperatures over the Antarctic Plateau.

ERA-Interim has an understandable problem in reproducing the observations from one of the six stations, Marambio. This station is located on an island close to and east of the northern limit of the Antarctic Peninsula. The topography of the region is not resolved by the assimilating model used by ERA-Interim, which treats the location as a sea point that generally has only partial ice cover. As a result, the background temperature lacks the interannual variation that occurs in the observations, as is evident in Figure 7.

The ERA-Interim background captures the observed interannual variability quite well for the other stations. This provides some confidence in the variability described by the background elsewhere around the continent, as the surface air temperature observations provide essentially independent validating data for the region, as discussed for the Arctic by Simmons and Poli (2015). Trends over the period are mixed for these stations, and smaller in magnitude than the warming trends that predominate in the Arctic. Maps of the ERA-Interim trend, which are similar for the background and for the analysis, nevertheless show regions of strong warming, most notably over central West Antarctica, for which discussion is given by Bromwich *et al.* (2013) using observations available intermittently from Byrd Station. A cooling trend offshore of East Antarctica is consistent with an increase of sea-ice concentration there over recent decades, which is discussed in IPCC (2013).

7.3. *Averages from 60°N to 60°S*

Figure 8 complements the information provided for the Arctic and Antarctic by presenting information for the region from 60°N to 60°S, in this case separated into the contributions

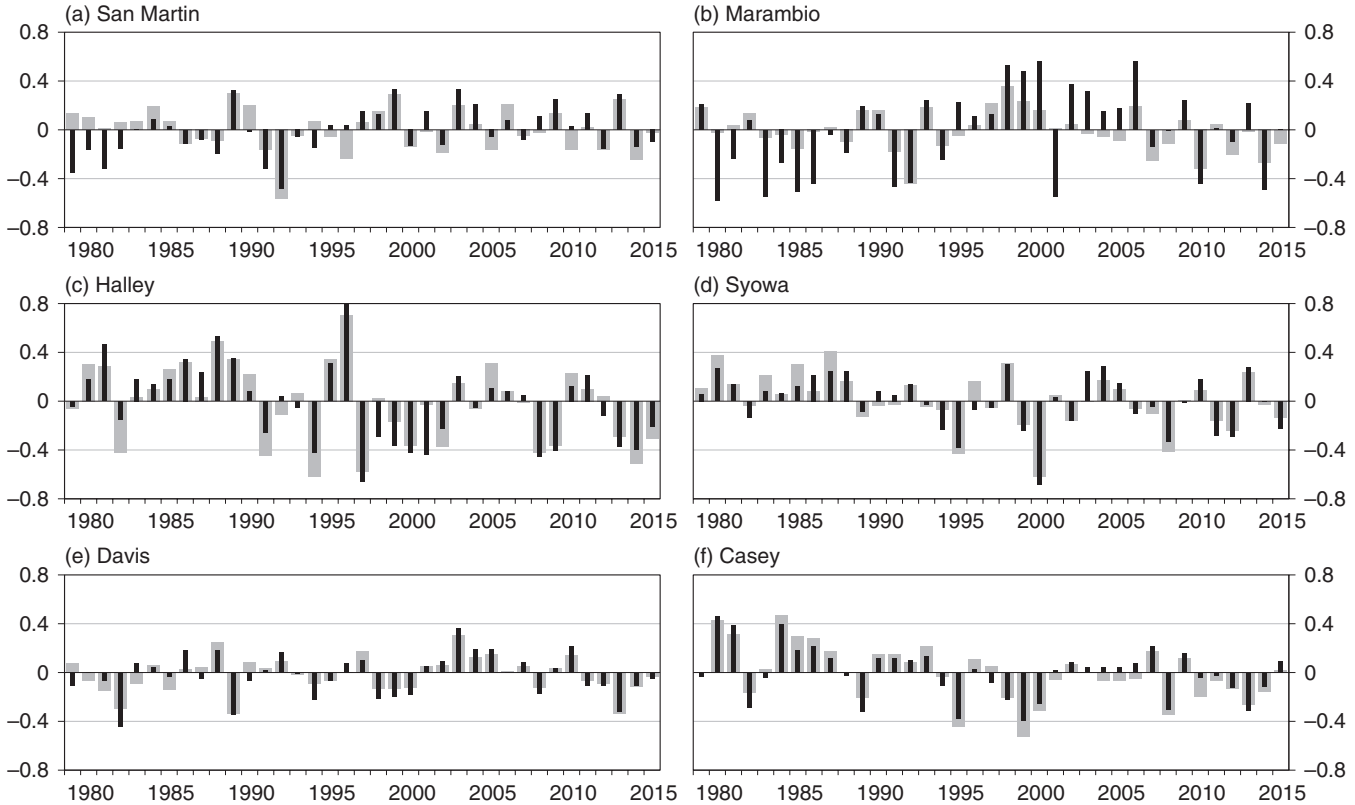

Figure 7. Annual-mean surface air temperature anomalies (°C) relative to 1981–2010, at (a) San Martin (68.1°S, 67.1°W), (b) Marambio (64.2°S, 56.7°W) (c) Halley (75.6°S, 26.7°W), (d) Syowa (69.0°S, 39.6°E), (e) Davis (68.6°S, 78.0°E) and (f) Casey (66.3°S, 110.5°E). Narrow, darker bars denote observed station values and broader, lighter bars denote corresponding ERA-Interim background forecast values.

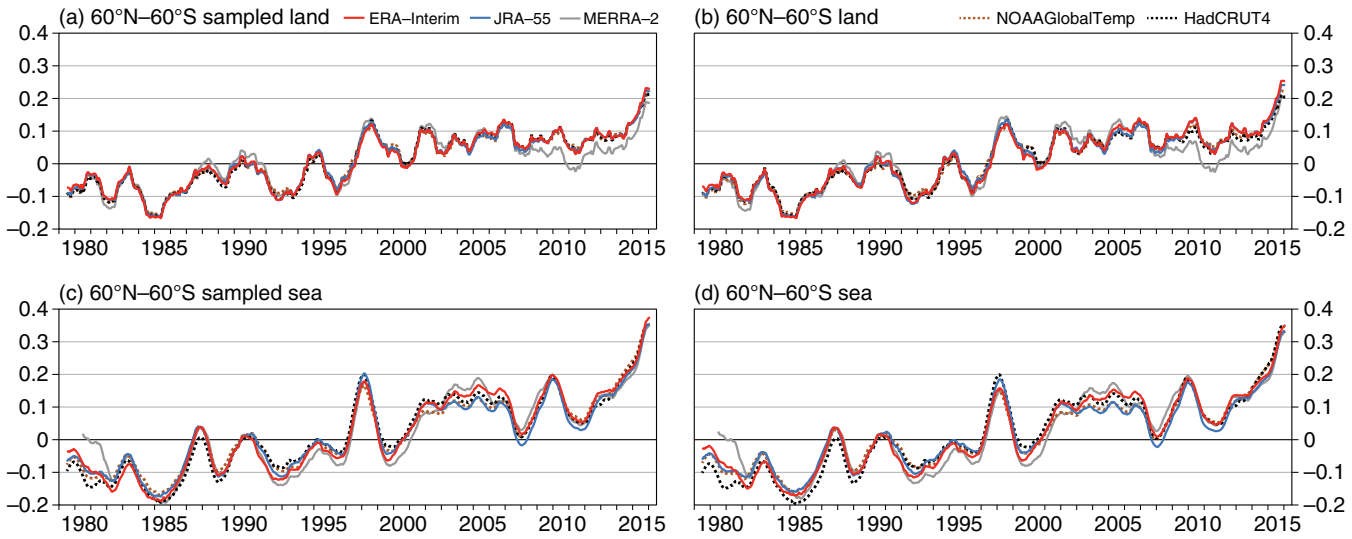

Figure 8. Contributions from (a, b) land and (c, d) sea to 12 month running-mean surface temperature anomalies relative to 1981–2010 (°C) based on data from January 1979 to July 2016, for the region from 60°N to 60°S. (a) and (c) are for HadCRUT4 (black, dotted), NOAAGlobalTemp (orange, dotted), ERA-Interim (red, solid), JRA-55 (blue, solid) and MERRA-2 (grey, solid) averaged over the sets of 5° grid boxes where all datasets provide values for land and sea respectively. (b) and (d) are for HadCRUT4 and NOAAGlobalTemp averaged over all grid boxes where each separately provides values, and for ERA-Interim, JRA-55 and MERRA-2 averaged over all land and sea.

from land and sea to the average temperature for the zone as a whole. The datasets are in clear agreement as to the warmth of both land and sea over recent months, particularly so for land, for which the major discrepancy occurs for MERRA-2. This contributes together with the polar regions to cause MERRA-2 to underestimate substantially the recent warming seen in all the other datasets considered in this article. The only other difference worthy of note for the land is the provision of slightly lower values since 2009 by HadCRUT4, and by implication CRUTEM4. This comes mainly from limitations in spatial coverage.

Differences are larger over sea than land. MERRA-2 agrees quite well with the other datasets from 2009 onwards, when it uses the same OSTIA SST analysis as ERA-Interim, but it also exhibits shifts associated with changes in SST analysis, which are discussed by Bosilovich *et al.* (2015). Otherwise, HadCRUT4 has a slightly larger trend than ERA-Interim, while the COBE SST and the corresponding JRA-55 marine air temperature analysis have a slightly smaller one. There is also a particular spread of marine values from 2003 to 2006, with ERA-Interim distinctly warmer than JRA-55 relative to their respective 1981–2010 means. Maps show the mean 2003–2006 differences in SST anomalies between ERA-Interim and JRA-55 to be geographically widespread and largely of the same sign. The values from NOAAGlobalTemp (and by implication GISTEMP) are also lower than those from ERA-Interim over this period. The differences in SST anomalies are also relatively large from 1979 to 1982.

Twelve-month running mean temperatures over sea continue to rise through to the end of the period shown in Figure 8. Although monthly anomalies peaked in December 2015, values for the first few months of 2016 were considerably higher than those for the corresponding months of 2015, or indeed for any other year. Expressed as a contribution to the complete 60°N–60°S average, these values are at least 0.2 °C higher than at the times of preceding El Niño events. This is in contrast to the situation in the tropical eastern Pacific, where peak El Niño temperatures (specifically averages for the region from 5°N to 5°S and 180°W to 80°W) were a little lower in the 2015/2016 event than in the 1997/1998 event. More-generally high SSTs, as well as the latest El Niño and low Arctic sea-ice cover, thus appear to contribute to the exceptional recent values of global-mean temperature.

7.4. *Contributions to global means*

Table 2 presents summary information for the globally complete ERA-Interim, JRA-55 and Had4_UAH_v2 datasets, and for

Table 2. Anomalies in global-mean surface temperature (°C) relative to 1981–2010 for the years 2005, 2006 and 2010–2015 from ERA-Interim, JRA-55, GISTEMP and Had4_UAH_v2, and contributions from the zonal bands 60°N–90°N, 60°N–60°S and 60°S–90°S.

Headed area	Dataset	2005	2006	2010	2011	2012	2013	2014	2015
Global	ERA-Interim	0.35	0.29	0.31	0.19	0.22	0.25	0.29	0.44
	JRA-55	0.27	0.22	0.29	0.18	0.21	0.25	0.30	0.43
	GISTEMP	0.26	0.20	0.28	0.18	0.21	0.23	0.32	0.44
	Had4_UAH_v2	0.29	0.23	0.32	0.20	0.22	0.25	0.32	0.45
60°N–90°N	ERA-Interim	0.09	0.07	0.08	0.09	0.08	0.05	0.08	0.07
	JRA-55	0.08	0.06	0.08	0.09	0.08	0.05	0.08	0.07
	GISTEMP	0.08	0.05	0.07	0.08	0.06	0.03	0.06	0.06
	Had4_UAH_v2	0.07	0.05	0.08	0.09	0.08	0.04	0.07	0.07
60°N–60°S	ERA-Interim	0.22	0.21	0.25	0.08	0.15	0.21	0.24	0.41
	JRA-55	0.18	0.16	0.23	0.06	0.13	0.20	0.23	0.39
	GISTEMP	0.18	0.16	0.23	0.08	0.15	0.19	0.25	0.41
	Had4_UAH_v2	0.20	0.18	0.24	0.08	0.14	0.18	0.24	0.40
60°S–90°S	ERA-Interim	0.04	0.02	−0.02	0.02	−0.01	−0.01	−0.03	−0.04
	JRA-55	0.02	0.01	−0.02	0.03	0.00	0.00	−0.01	−0.03
	GISTEMP	0.02	−0.01	0.00	0.02	−0.01	0.02	0.00	−0.04
	Had4_UAH_v2	0.02	0.00	0.00	0.03	0.00	0.02	0.01	−0.02

For GISTEMP, the sum of the regional contributions differs from the global value by slightly more than can be explained by rounding for some years. This stems from the differences in processing global and regional GISTEMP data noted in section 2.

GISTEMP, which has few missing values. It shows annual averages for the years 2005, 2006 and 2010–2015 of the global-mean surface temperature and of the contributions to the global mean from the regions 60°–90°N, 60°N–60°S and 60°S–90°S. All these years have temperatures that are globally above the 1981–2010 average. The four datasets are in good agreement for 2010–2015, during which the contribution from the relatively warm Arctic to the global mean varies from around 50% in 2011 to 15% in 2015. The contribution from the Antarctic is small.

More difference is seen in 2005 and 2006, when the anomaly for ERA-Interim is larger than for any of the other datasets considered, as shown already in Figure 1. For these years, a significant contribution to the differences between ERA-Interim on the one hand and JRA-55 and GISTEMP on the other stems from the SST differences reported above. ERA-Interim also has larger positive anomalies in both the Arctic and the Antarctic at the time, as can be seen also in Figure 6. Although some additional discussion is given in section 10, further investigation seeking to establish which of the datasets is the most trustworthy in each of the domains is beyond the scope of this study.

8. Longer-term data records

The preceding discussion has focused on the period since 1979, and the latest warm spell in particular. ERA-Interim does not go back earlier. JRA-55 runs from 1958, but neither reanalysis comes close to matching the length of record of the established conventional datasets. Century-scale reanalyses that assimilate surface pressure and in some cases wind observations but no other meteorological data are available, but their agreement over land with CRUTEM4 has been shown to be poorer on annual and longer time-scales than that of atmospheric model simulations using similar SST analyses and external forcings (Hersbach *et al.*, 2015).

The surface warming of the atmosphere since 1979 is substantially larger than the differences between the estimates provided by the various datasets examined, for both global and European averages. Appeal may thus be made to the behaviour of the longer-term datasets to establish which statements about warm extremes made on the basis of ERA-Interim and other datasets available only for the past few decades can be expected to hold over a much longer period. The HadCRUT4 dataset is particularly useful for this purpose, as it extends the farthest back in time, to 1850, and gives information on uncertainty through its ensemble of 100 possible realizations.

Figure 9 presents the time series to the present day of 12 month running averages of global- and European-mean surface temperatures for the HadCRUT4 median from 1850 and NOAAGlobalTemp from 1880. Various time series of differences are also shown, relating to the HadCRUT4 ensemble and

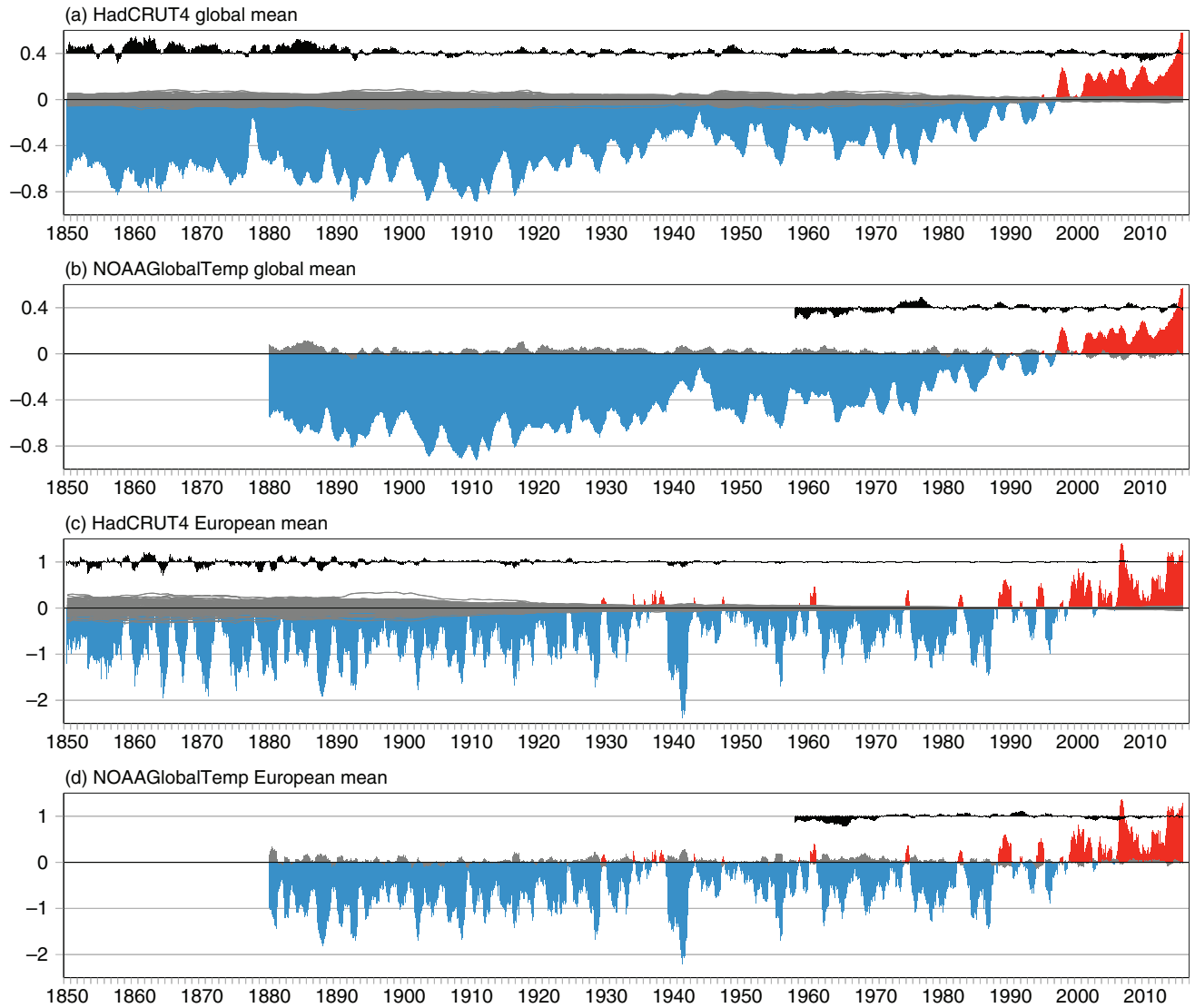

Figure 9. Twelve-month running means of anomalies in (a, b) global-average and (c, d) European-average surface temperatures (°C) relative to 1981–2010, for the full periods of record to July 2016 from (a, c) HadCRUT4 and (b, d) NOAAGlobalTemp. The dark grey lines in (a) and (c) denote the differences between the values of the 100 HadCRUT4 ensemble members and the HadCRUT4 median. In (b) and (d) the differences between NOAAGlobalTemp and GISTEMP values are shown in grey. The differences between HadCRUT4 and Had4_krig_v2 are shown in black in (a) and (c), shifted by 0.4 and 1 °C respectively for clarity, and differences between GISTEMP and JRA-55 are shown in a similar manner in (b) and (d).

Had4_krig_v2, and to GISTEMP and JRA-55. All maxima prior to 1979 in all displayed datasets lie well below the maxima that occur from the late 1980s onwards. The same is true of the model simulations and century-scale reanalyses examined by Hersbach *et al.* (2015). This makes it highly probable that statements made concerning the extreme warmth of recent months based on ERA-Interim and other limited-duration data records apply for the whole period since 1850, at least as regards 12 month averages.

NOAAGlobalTemp is generally similar to HadCRUT4 over the period of common record from January 1880 onwards. Both datasets show a spell of anomalously low global-mean temperature early in the twentieth century, and relatively warm conditions in the first half of the 1940s, notwithstanding markedly below-average European temperatures during the early years of the Second World War. Relative to their respective 1981–2010 means, NOAAGlobalTemp is on average colder than the HadCRUT4 median for earlier years, and GISTEMP is colder still. The mean differences between 1880–1980 and 1981–2010 averages are −0.49 °C for HadCRUT4, −0.50 °C for Had4_krig_v2, −0.54 °C for NOAAGlobalTemp and −0.57 °C for GISTEMP. The global-mean temperature anomalies averaged for 1911–1940 and 1941–1970 from NOAAGlobalTemp and GISTEMP are in each case lower than the lowest corresponding 30 year average from the HadCRUT4 ensemble. Differences are much smaller for 1881–1910.

The global means from the two datasets with increased global coverage, GISTEMP and Had4_krig_v2, are mostly a little lower respectively than NOAAGlobalTemp and HadCRUT4. Differences are larger in the earlier years when observational coverage is poorer, especially in the case of Had4_krig_v2 and HadCRUT4. However, in contrast to the quite similar changes they bring to the representation of the short-term variability of recent years, GISTEMP and Had4_krig_v2 remain closest in earlier years to the datasets that use the same SST analyses, respectively NOAAGlobalTemp and HadCRUT4.

Figure 10 illustrates this in maps of averages for the periods 1881–1910, 1911–1940 and 1941–1970 relative to 1981–2010. As data coverage is poorer for the earlier periods, values for a grid square are plotted here if there is at least 80% rather than 90% data availability for each period. The anomalies from NOAA-GlobalTemp and GISTEMP for 1911–1940 and 1941–1970 can be seen to be generally slightly lower over the oceans than those from HadCRUT4 and Had4_krig_v2. For both periods, NOAA-GlobalTemp is lower than HadCRUT4 when averaged over either

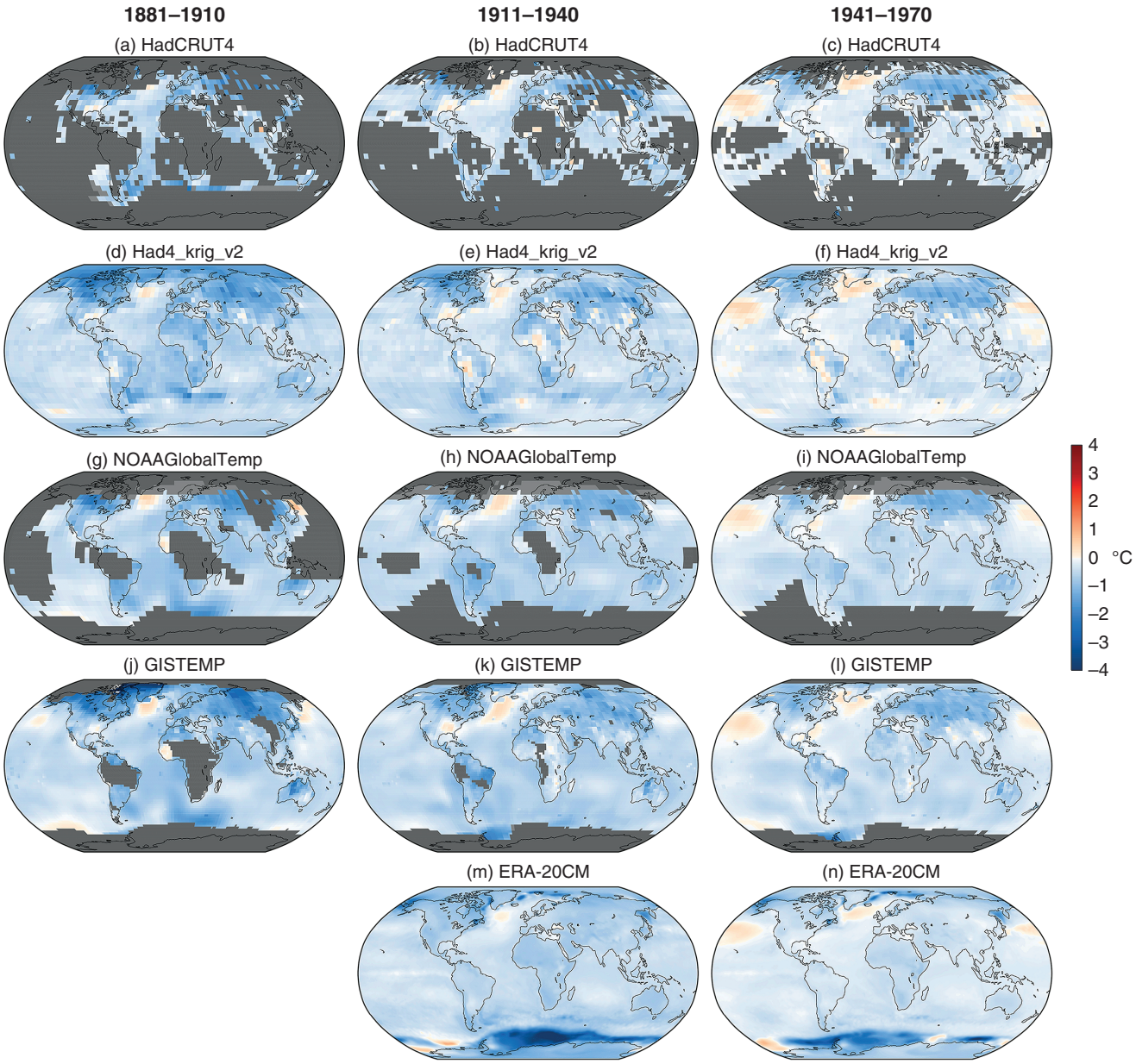

Figure 10. Surface temperature anomalies (°C) relative to 1981–2010 from (a, b, c) HadCRUT4, (d, e, f) Had4_krig_v2, (g, h, i) NOAAGlobalTemp and (j, k, l) GISTEMP for (j) 1881–1910, (k) 1911–1940 and (l) 1941–1970, and from the ERA-20CM ensemble mean for (m) 1911–1940 and (n) 1941–1970. Grid boxes where values are missing are coloured grey. Lighter grey colouring indicates boxes that would have had values had maps been presented as anomalies relative to the datasets' standard reference periods. In (m) and (n), two small regions around Antarctica where negative anomalies exceed 4 °C have been shaded at the 4 °C level.

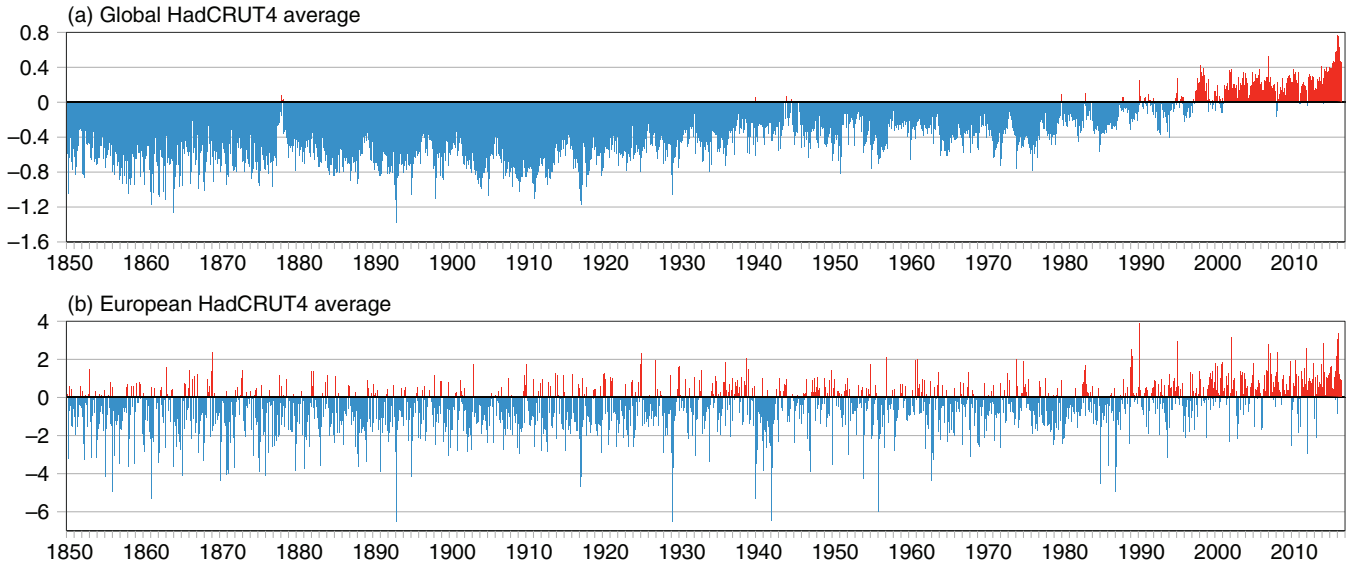

Figure 11. Monthly anomalies in (a) global-average and (b) European-average surface temperature (°C) relative to 1981–2010, from HadCRUT4 for January 1850 to July 2016.

the land or the sea areas where both provide data, but differences are larger over sea, even though the anomalies are larger over land. The datasets give more similar averages for 1881–1910.

GISTEMP and Had4_krig_v2 also differ quite substantially in the earlier years over the regions where they perform substantial infilling. Figure 10 shows pronounced differences in the infilling over Africa and South America, with Had4_krig_v2 predominantly lower than GISTEMP over Africa but higher over South America. These two datasets also differ over the Atlantic sector of the Arctic Ocean for 1911–1940.

Figure 10 includes corresponding plots for 1911–1940 and 1941–1970 from the mean of the ten-member ERA-20CM ensemble of atmospheric model integrations, which used prescribed time-varying SSTs, sea-ice distributions, solar radiative forcings and radiatively active trace gases and aerosols (Hersbach *et al.*, 2015). The ensemble averaging produces somewhat smoother temperature distributions than those of the observational datasets, but the ERA-20CM temperature anomalies over land are generally smaller, more so for 1941–1970. This is shown more fully in the comparison with CRUTEM4 presented by Hersbach *et al.* (2015). Over ice-free sea the ERA-20CM temperature anomalies are constrained by the geographically complete HadISST2 (version 2.1.0.0) analyses. As might be expected, the anomalies are closer to those of the HadSST3-based HadCRUT4 and Had4_krig_v2 datasets than those of the ERSSTv4-based NOAAGlobalTemp and GISTEMP datasets. ERA-20CM has distinctly lower temperatures in earlier than later decades over regions where HADISST2 indicates greater sea-ice concentrations in the earlier years. This includes the Antarctic, notwithstanding the observed increase in sea-ice since the 1970s. Such features in the temperature field are at least qualitatively in line with expectations, but are not reproduced by the methods of geographical extension used in deriving GISTEMP and Had4_krig_v2.

Europe has reasonably complete data coverage throughout the period of record of HadCRUT4. Figure 9 shows that the differences between HadCRUT4 and Had4_krig_v2 are accordingly small. This is particularly so from around 1900 onwards, although differences can be seen to be subsequently a little larger at the times of the two World Wars. However, there are other sources of uncertainty which lead to a quite considerable spread among the HadCRUT4 ensemble for the nineteenth century and early decades of the twentieth century. A single member stands out for almost 30 years from around 1890 in having a much higher European-average temperature anomaly. Over this period, the error model used for CRUTEM4 (Brohan *et al.*, 2006) persistently generates extreme values for two European grid squares of one ensemble member.

The differences between the JRA-55 and GISTEMP global-mean anomalies shown in Figure 9 are larger before 1979 than afterwards, and the differences in the anomalies for Europe are relatively large in the 1960s. This is probably due both to the generally poorer global observing system available for reanalysis prior to 1979 and to specific gaps in coverage of the surface synoptic observations used by JRA-55 for the 1960s. The pre-1979 data used by JRA-55 were largely based on the collection of data made earlier for ERA-40, which lacked surface synoptic data from several countries prior to 1967, including European ones (Simmons *et al.*, 2004).

Time series of monthly values since 1850 from HadCRUT4 are shown in Figure 11. Only median values are shown, for clarity of display. For the global average it is again highly probable that statements concerning the extremity of the warmer periods of the last four decades in fact apply for the record back to 1850. However, it is evident for Europe that individual months can be relatively warm right back to 1850, even if the frequency of warm months is higher in recent decades, and the warmest few months occur over the past 30 years.

The monthly extremes from HadCRUT4 are larger than those from NOAAGlobalTemp but smaller than those from GISTEMP. For example, the highest monthly European temperature anomaly prior to 1989 is a little below 2 °C in NOAAGlobalTemp, whereas HadCRUT4 identifies five warmer months between 1880 and 1988, and GISTEMP ten. HadCRUT4's highest pre-1989 value, 2.4 °C for February 1869, is nevertheless larger than any of the 1880–1980 GISTEMP values. These figures help place in context the recent positive European anomalies seen in Figure 5, which are 3.2 °C for December 2015 and 3.9 °C for February 2016 from ERA-Interim. The corresponding anomalies from JRA-55 are 3.3 °C and 3.7 °C. HadCRUT4 gives smaller values, 3.0 °C and 3.4 °C, but its lower spatial resolution has to be taken into account: ERA-Interim anomalies are reduced to 2.9 °C and 3.4 °C respectively if an equivalent of the 5° × 5° HadCRUT4 dataset is first constructed from ERA-Interim, and then averaged over European land areas following what was done for HadCRUT4.

9. Atmospheric energy and upper-air temperature

Variability over a range of time-scales makes global-mean surface temperature challenging to use as a metric of climate change, whether as a target of the Paris Agreement or as the measure of response to a doubling of atmospheric carbon dioxide in the concepts of equilibrium and transient climate sensitivity (IPCC, 2007, 2013). Surface air temperature is dependent on fluctuations in tropical SST and winter sea-ice cover among other

factors, but the vertical extent of the associated perturbations in upper-air temperature varies considerably between the two, and a particular warm event may be more or less extreme at the surface than in the upper troposphere, as illustrated in this section. More generally, Davy and Esau (2016) discuss how the response of surface air temperature to a forcing may very regionally, and be determined by the depth of the planetary boundary layer in regions that are cold and dry.

Atmospheric energy provides another metric of global change, for which not only the different distributions of surface and upper-air temperature but also moisture content has to be taken into account when comparing its variations with those of global-mean surface temperature. Peterson *et al.* (2011) discuss this for the near-surface atmosphere over land, showing that thermal energy (or enthalpy) and latent energy (or latent heat) generally but not always change in concert. They note also the large disparity in energy content between the lowest 2 m of the atmosphere and the upper 2 m of the oceans.

The oceans as a whole have been estimated to have absorbed about 93% of the increase in energy of the climate system between 1971 and 2010, with melting ice and warming land accounting for much of the remainder; the estimated increase in atmospheric energy accounts for only about 1% of the net increase in energy of the system (Box 3.1 of Rhein *et al.*, 2013). This estimate for the atmosphere was based on the microwave temperature sounding record of Mears and Wentz (2009), which like the UAH record has limited vertical resolution. It also assumed a fractional increase in water vapour content as temperature increases, and neglected changes in potential and kinetic energy.

Reanalysis data enable a complete calculation of changes over time in atmospheric energy, though subject to provisos as to the temporal consistency of its estimates of the underlying state variables. Figure 12 shows time series (up to August 2016) of the anomalies relative to 1981–2010 in total energy and in its dominant thermal and latent energy components, based on ERA-Interim data, for which Berrisford *et al.* (2011) discuss related global budgets. The corresponding time series of global-mean surface air temperature is included for comparison. In addition to the dataset adjustments described previously for surface air temperature, the latent energy and its contribution to total energy have been increased by 3% after 1991 to compensate for a marked shift in ERA-Interim values relative to values both from ERA-20CM and from microwave imagery (Hersbach *et al.*, 2015; Poli *et al.*, 2016). The chosen units make the values of the anomalies in thermal energy close to those for the global mass-weighted average atmospheric temperature in °C. The 2 TW average rate of change in total energy from 1979 to 2010 estimated by Rhein *et al.* (2013) is equivalent to a shift of about 0.4 over this period in the units used in Figure 12.

The anomalous spells seen in the various time series shown in Figure 12 are generally common to all series, but their magnitudes vary considerably. The anomalies in winter surface air temperature tend to be associated with relatively shallow structures in the vertical, and thus do not show as prominently in thermal energy. In contrast, variations in surface air temperature in the Tropics tend to be associated with larger variations in upper-tropospheric temperature, and thus with more pronounced features in thermal energy. Anomalies in latent energy are most pronounced in the Tropics and Subtropics. The El Niño events of 1997/1998, 2009/2010 and 2015/2016 are thus much more marked for total energy than for surface air temperature. A tendency for tropospheric cooling due to the 1982 volcanic eruption of El Chichón counters warming due to the 1982/83 El Niño.

Broad peaks in latent energy occur during the 1997/1998, 2009/2010 and 2015/2016 El Niño events. Latent energy can be sustained at anomalous levels in such events because of the greater capacity of an anomalously warm atmosphere to carry water vapour, but is converted to thermal energy in the declining phase of events as anomalous amounts of moisture are removed from the atmosphere by increased precipitation. ERA-Interim's representation of the latter over land for the 1997/1998 and 2009/2010 events has been shown to agree well with the independent analyses of precipitation produced by the Global Precipitation Climatology Centre (Simmons *et al.*, 2014). This is found to be the case also for the latest event.

Variations in latent energy more generally tend to lead variations in thermal energy, in that correlations between their time series shown in Figure 12 are 75% and 72% with latent energy leading thermal energy by 1 and 2 months respectively, and 71% and 67% with latent energy lagging by 1 and 2 months, using the period from March 1979 to June 2016 for thermal energy. The zero-lag correlation between the time series is 77%. As is the case for surface air temperature, the annual-mean latent energy for 2015 is higher than for any previous 12 month period, whereas the annual-mean thermal energy for 2015 is lower than for both 1998 and 2010. The 12 month mean thermal energy first exceeds its earlier peak value in the mean for the 12 months that end in February 2016. Its value to August 2016, 0.51×10^7 J m^{-2} relative to the 1981–2010 average, is the largest on record, and well in excess of the previous peak values of 0.32×10^7 J m^{-2} for the calendar year 1998 and 0.35×10^7 J m^{-2} for the period from November 2009 to October 2010. The largest anomaly in 12 month mean latent energy, 0.35×10^7 J m^{-2}, is also reached for the period to August 2016.

Figure 13 shows time series of monthly anomalies in temperature at 700, 500 and 300 hPa from ERA-Interim and the differences between JRA-55 and ERA-Interim. The larger

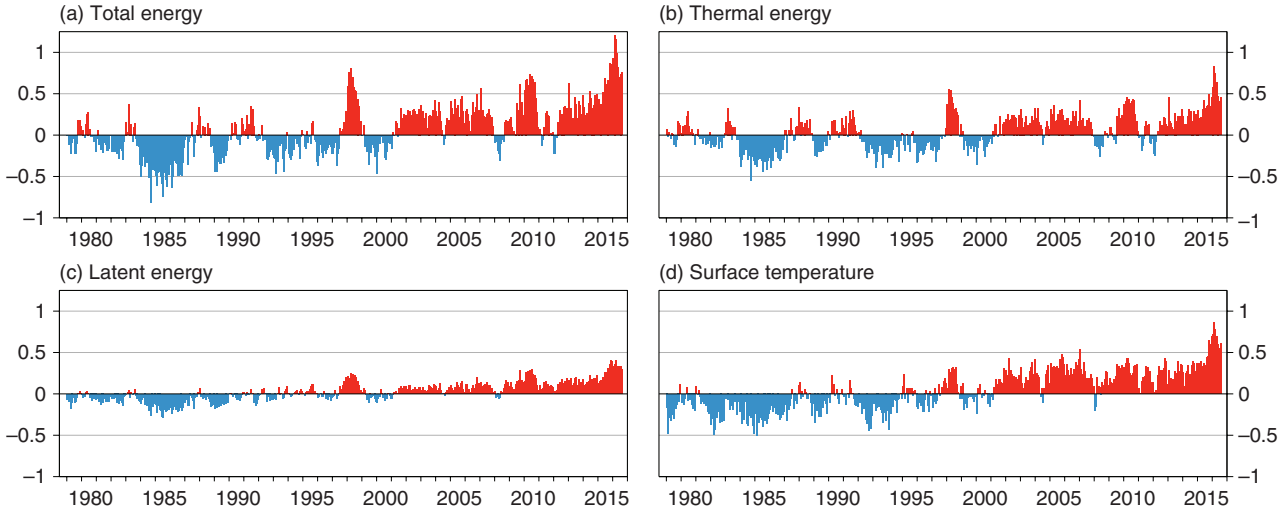

Figure 12. Monthly anomalies of atmospheric (a) total energy, (b) thermal energy and (c) latent energy expressed as average values per unit area of the Earth's surface (10^7 J m^{-2}), and of (d) global-mean surface air temperature (°C), relative to 1981–2010, from ERA-Interim for January 1979 to August 2016.

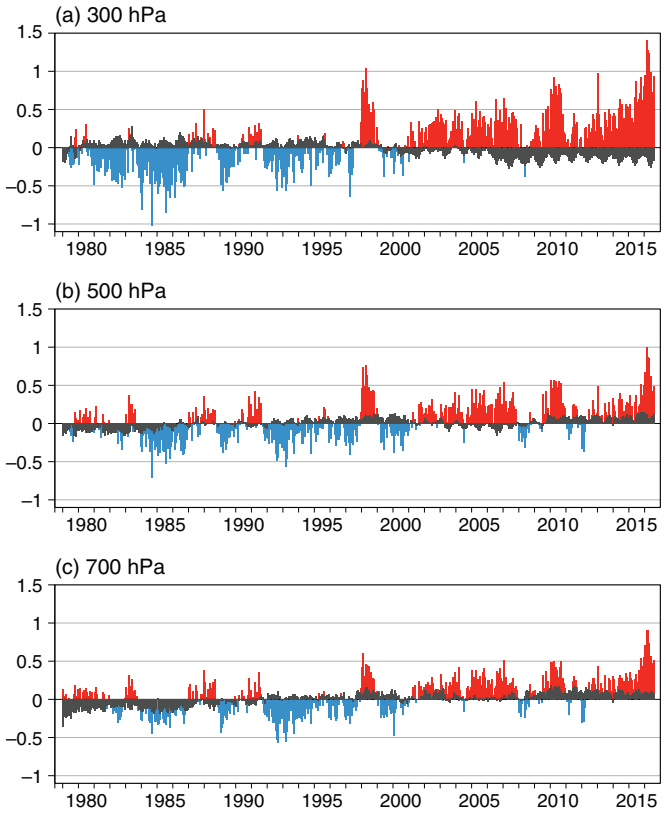

Figure 13. Monthly anomalies in globally averaged temperature (°C) relative to 1981–1990 at (a) 300 hPa, (b) 500 hPa and (c) 700 hPa, from ERA-Interim for January 1979 to August 2016, denoted by red and blue bars. Grey bars denote differences between JRA-55 and ERA-Interim.

amplitudes at higher levels of the perturbations associated with El Niño events are evident. Again in contrast to the situation for surface air temperature, recent 500 and 300 hPa temperature anomalies exceed their previous maximum value, reached in 1998, only for 2 or 3 months, from February 2016. This is the case for both ERA-Interim and JRA-55.

Although the two reanalyses tell essentially the same story, interpretation of detail in their time series is hampered by differences between the two. ERA-Interim exhibits a weaker overall warming trend over the period at 700 and 500 hPa; reasons why it is thought to underestimate trends in the lower and middle troposphere are discussed by Simmons *et al.* (2014). Conversely, it is JRA-55 that exhibits the weaker warming trend at 300 hPa. This comes in particular from a drop in temperature in July 2006, when assimilation of substantial additional information on upper-tropospheric and stratospheric temperatures from GPS radio occultation starts to reduce a positive bias that stems from the background model (Kobayashi *et al.*, 2015).

Figure 14 shows maps of the temperature anomalies at 700, 500 and 300 hPa for December 2015 and January and February 2016, from ERA-Interim. The surface air temperature anomaly over the tropical Pacific Ocean reached peak amplitude in November 2015 and was slowly declining over the subsequent 3 months, as shown earlier in Figure 5. An extensive warm anomaly over the eastern tropical and sub-tropical Pacific can be seen at 300 hPa in December, and warming of this region continued into January at all levels shown. Warmth was more widespread geographically by February. This can occur both directly through advection of sensible heat and through advection and remote release of latent heat. Radiative effects may also play a role. The shallower nature of the middle- and high-latitude anomalies that are predominant in the surface maps is also evident in Figure 14.

10. Concluding discussion

The latest versions of several well-established conventional datasets and two recent reanalyses have been shown to agree well in their depiction of the net warming that has taken place at the Earth's surface over the past three to four decades. They agree also on the general character of the variability that has occurred over this period, and on the extremity of global warmth that has occurred over the past 12 months as a strong El Niño, more-generally high SSTs and low Arctic sea-ice cover have raised atmospheric temperatures during the boreal winter of 2015/2016. The versions of the conventional datasets that infill data where they lack direct observational sources agree better with the reanalyses. Giving weight to these, the evidence indicates that, based on norms for 1981–2010, 12 month running averages of global-mean surface temperature anomalies have reached a little over 0.6 °C during the latest warm spell, and monthly anomalies peaked in the range 0.8–0.9 °C early in 2016.

Placing these numbers in the context of the Paris Agreement requires that an estimate be made of the extent to which the 1981–2010 norms for temperature lie above the global temperature for a pre-industrial period that is not precisely defined in the Agreement. There are also issues of sparser data coverage and poorer data quality in earlier years. Uncertainty in the amount of warming from the start of the Industrial Revolution, around the middle of the eighteenth century, to the final decades of the twentieth century is considerably larger than uncertainty in the amount of warming over the past three to four decades. Taking the warming since the latter half of the nineteenth century from the conventional datasets discussed in section 8 and adding 0.1 °C for earlier warming based on evidence presented by Masson-Delmotte *et al.* (2013) suggests a value 0.7 °C for the warming from the pre-industrial to 1981–2010, with a two-standard-deviation uncertainty range upwards of ±0.1 °C (Morice *et al.*, 2012). This is consistent with the range 0.55–0.75 °C derived by Hawkins *et al.* (2016; personal communication) for the warming from the pre-industrial period 1720–1800 to 1986–2005. In turn this suggests that global-mean surface temperatures have been close to or more than 1 °C above the mid-1700s level for the past two years, and that they peaked at around 1.5 °C above in February 2016. It is salutary to note from Figure 3 that on the same basis it was probably only in 1998 that the 1 °C level was first reached. This was in the latter stages of a somewhat stronger El Niño event than the latest one, but one that was accompanied by below- rather than above-average temperatures in the Arctic (Simmons and Poli, 2015) and by sea-surface temperatures that were on average lower away from the tropical eastern Pacific.

What constitutes damaging climate change cannot be encapsulated in the value of a single metric such as global-mean surface temperature, even if the latter arguably provides the best single quantity for which to express an overall target (Knutti *et al.*, 2015). Nevertheless, whatever metric is used in a particular case, the critical value or range of values must be an absolute one, even if imperfectly known, not one relative to a pre-industrial level of uncertain value that is likely to change as data, modelling and understanding are refined. Targets for limiting future change, to be achieved by limiting anthropogenic disruption of the climate system, would be better framed and monitored in a global stocktake in terms of change relative to a recent period over which the system has been comparatively well observed. This does not mean that work to improve estimation and understanding of change from the pre-industrial to the recent past is not needed, as it serves purposes that include evaluating climate models and determining responsibilities for past change and its impacts. Splitting the calculation of change into two parts would allow the conventional approaches to be improved for application to recent decades, for example to use SST analyses that draw on satellite as well as *in situ* data, or to use satellite data to help fill gaps over land and sea-ice, as already done by Cowtan and Way (2014) in their hybrid approach to spatially extending HadCRUT4.

The absence of infilling and spatial smoothing in HadCRUT4 has the advantage of identifying grid squares where data are most directly tied to conventional climate observations and

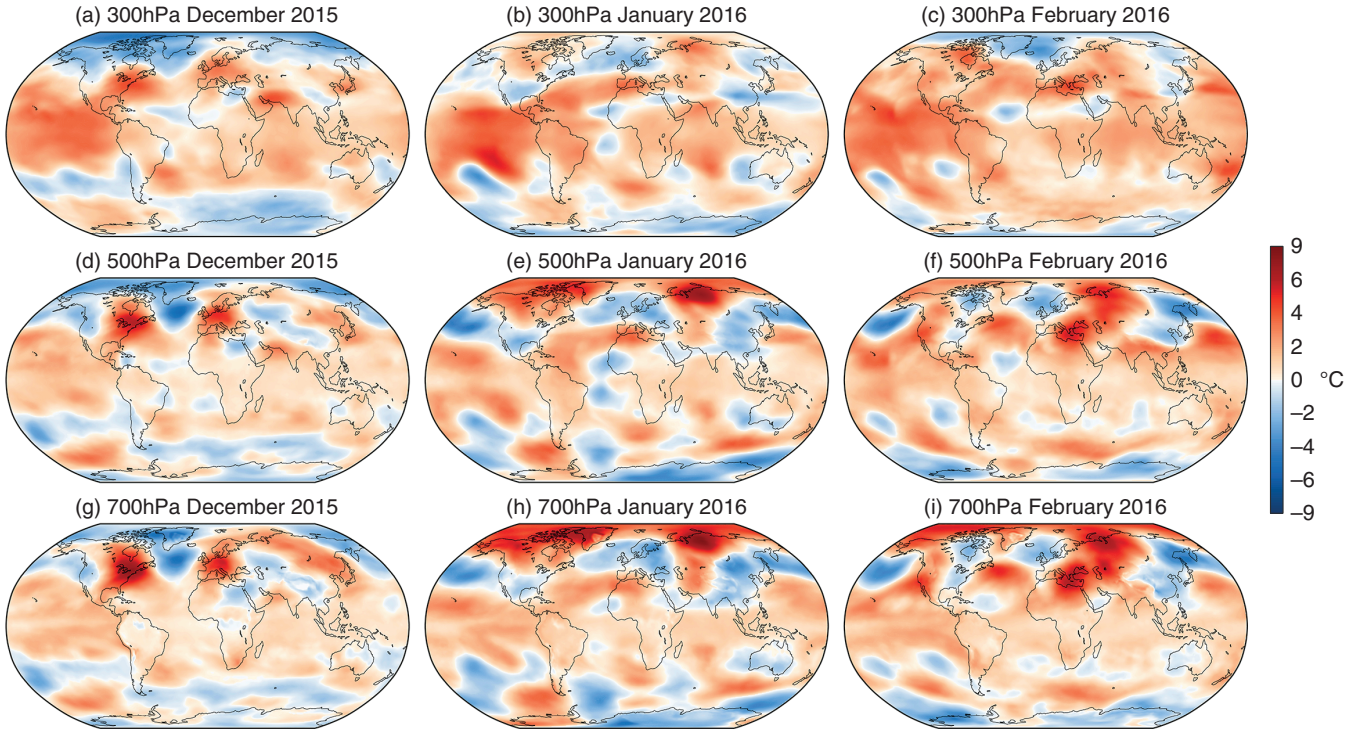

Figure 14. Anomalies in monthly average ERA-Interim temperatures (°C) relative to 1981–1990, at (a, b, c) 300 hPa, (d, e, f) 500 hPa and (g, h, i) 700 hPa, for (a, d, g) December 2015, (b, e, h) January 2016 and (c, f, i) February 2016.

thus particularly suitable for cross-checking with the results of reanalysis. However, it does have the downside of making the HadCRUT4 median prone to be an outlier, separated from the consensus of other datasets by more than might be expected from the spread of the underlying ensemble of realizations. This is not unexpected when it applies to the monthly variations examined here for recent years, as the HadCRUT4 ensemble samples uncertainty primarily on multi-decadal time-scales. However, if all data are expressed relative to 1981–2010, the ensemble also does not have the spread to encompass the values from GISTEMP and NOAAGlobalTemp in averages for 1911–1940 and 1941–1970, echoing a similar AR5 finding for the decadal averages from previous versions of the datasets (Hartmann *et al.*, 2013). Differences in SST analyses are the main contributor. It is important to recall in this regard that GISTEMP and NOAAGlobalTemp use the same SST analyses, and should not be regarded as independent datasets when combined or compared with HadCRUT4. Also, differences between the infillings provided by GISTEMP and Had4_krig_v2 and the absence of features evidently linked to sea-ice changes indicate uncertainties in these datasets for the earlier periods.

The present results provide more evidence of the value of reanalysis as an important complement to the conventional analyses for depicting variability and change in surface temperature. Reanalysis exploits the richness of the observing system that has been in place over recent decades. ERA-Interim and JRA-55 have set a standard for the period in the sense that newer versions of the conventional datasets, and those datasets that extend data coverage spatially, are closer to the two reanalyses than are earlier versions and those datasets with more limited coverage. This is seen in particular in the estimates of the global temperature trend over the period 1998–2012 discussed only for the conventional datasets in AR5. Reanalysis also provides data with higher spatial and temporal resolution, and linked information on related variables such as surface humidity and precipitation (Simmons *et al.*, 2010, 2014) and atmospheric energy as discussed in this article. The conventional datasets still provide the primary source of information for earlier decades, for which data recovery and improved data assimilation are needed to advance the contribution of centennial-scale reanalysis.

Nevertheless, use of reanalysis to monitor recent and future change requires a careful, comparative and selective approach. More is needed than simply displaying multiple time series from all or many of the available products and judging uncertainty from the spread of values. Some reanalyses are more fit for purpose than others for a particular application such as surface temperature trends, whereas others may be competitive for other applications. Although newer reanalyses are expected generally to perform better than older ones, ERA-Interim and JRA-55 have been shown here to perform more consistently with each other and with the conventional surface temperature datasets than the newer MERRA-2 does. Moreover, it has been necessary even for ERA-Interim and JRA-55 to combine background fields over sea and analysis data over land, and to make adjustments for inconsistencies in SST analysis for ERA-Interim, in order to achieve the reported level of agreement.

The need for high quality SST analyses is common to the conventional datasets and the reanalyses, although in the longer term there is the prospect for reanalysis to use direct assimilation of SST observations in systems that couple atmosphere, ocean and sea ice (Dee *et al.*, 2014; Laloyaux *et al.*, 2016). The most significant difference in SST anomalies between ERA-Interim and JRA-55 occurs between 2003 and 2006, when the COBE SST used by JRA-55, in common with the ERSSTv4 used by GISTEMP and NOAAGlobalTemp, gives lower values than the operational SST analysis of the US National Centers for Environmental Prediction (Gemmill *et al.*, 2007) used by ERA-Interim. The difference is just sufficient by itself to account for 2014 being a warmer year than 2005 in JRA-55 but not ERA-Interim. ERA5, the replacement for ERA-Interim, will not resolve the difference, as all the ten HadISST2 realizations it will use prior to 2007 in its ensemble data assimilation differ little from the ERA-Interim SST over the period in question (Hirahara *et al.*, 2016). Moreover, the year 2010 is warmer than 2014 in both the spatially extended Had4_UAH_v2 dataset and ERA-Interim. Statements that 2014 is ranked the second warmest year globally after 2015 are quite often made in the media, but do not have an especially strong basis.

ERA-Interim is also relatively warm in the global mean for 2005 because it has a larger positive anomaly in the Antarctic than the other datasets, including contributions in winter and spring

from regions with anomalously low sea-ice concentrations. Direct observational data on surface air temperature are sparse for the Antarctic, but none of the datasets examined provides evidence of net warming south of 60°S since 1979, a period during which sea-ice extent increased a little. Under-sampling of this region when estimating global means from the conventional datasets thus tends to compensate for under-sampling of the Arctic, where warming has been greater than elsewhere. The agreement between datasets concerning the global temperature trend since 1979 may thus be partly fortuitous, as evidence points to warming of the Antarctic over the preceding two decades and model projections are for future warming, although confidence is low in several aspects of observation, modelling and understanding for the region (IPCC, 2013).

The comparisons presented here support the current use of ERA-Interim as the primary source of information for the temperature summaries issued early each month by the Copernicus Climate Change Service (at http://climate.copernicus.eu), although applications will often need to consider the range of reasonable estimates provided by a set of datasets such as studied here. The Copernicus summaries extend time series and update maps of the type shown in this article. They are based on provisional ERA-Interim analyses for the preceding month, which can be cross-checked with promptly available JRA-55 analyses and with some of the published national information that the reanalyses help place in regional and global contexts. Observational data supply for the conventional datasets is slower, and results from them are normally available 2–4 weeks after the end of the month; these datasets serve to provide longer-term context and indications of uncertainties. Complete ERA-Interim data are released only after a delay of around 2 months, following more comprehensive checks. This allows short periods to be rerun if remediable problems are detected.

Acknowledgements

This study has been partially funded by the Copernicus Climate Change Service. ECMWF implements this Service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. Paul Berrisford's contribution has been supported by NCAS-Climate. The three reviewers of this paper are thanked for many helpful comments on the original text and figures.

Appendix: Adjustment of ERA-Interim temperatures

Global- and European-mean monthly anomalies in 2 m temperatures relative to 1981–2010 from ERA-Interim, adjusted over sea as described in section 2 and including provisional values for the latest 2 months, are provided for download in the monthly temperature summaries published at http://climate.copernicus.eu

Adjusted fields of monthly-mean 2 m ERA-Interim temperatures can be calculated straightforwardly from archived values of several fields: the invariant fractional land–sea mask (*lsm*), and the monthly mean 2 m analysed and background forecast temperatures ($T_{2\mathrm{man}}$ and $T_{2\mathrm{mfc}}$, in K) and fractional sea-ice cover (*sic*). The adjusted 2 m temperature is:

$$lsm \times T_{2\mathrm{man}} + (1 - lsm) \times (T_{2\mathrm{mfc}} - sst_{\mathrm{adj}})$$

where

$$sst_{\mathrm{adj}} = 0 \text{ from January 2002 onwards}$$

and

$$sst_{\mathrm{adj}} = 0.1 \times (1 - sic) \text{ prior to January 2002.}$$

The archived monthly mean SSTs (in K) are similarly adjusted by subtracting sst_{adj}.

Fields apart from $T_{2\mathrm{mfc}}$ can be downloaded directly from the public data interface at http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ choosing either the 'Invariant' or the 'Monthly Means of Daily Means' option. $T_{2\mathrm{mfc}}$ can be calculated by averaging 6 and 12 h forecasts starting from 0000 and 1200 UTC, which are directly downloadable choosing the 'Synoptic Monthly Means' option. Alternatively $T_{2\mathrm{mfc}}$ can be downloaded from the 'Monthly Means of Daily Means' archive using script access to the archive. User support for script access can be accessed via http://copernicus-support.ecmwf.int

References

Berrisford P, Dee D, Poli P, Brugge R, Fielding K, Fuentes M, Kållberg P, Kobayashi S, Uppala S, Simmons A. 2011. 'The ERA-Interim archive. Version 2.0', ERA report Series No. 1. ECMWF: Reading, UK.

Bosilovich MG, Akella S, Coy L, Cullather R, Draper C, Gelaro R, Kovach R, Liu Q, Molod A, Norris P, Wargan K, Chao W, Reichle R, Takacs L, Vikhliaev Y, Bloom S, Collow A, Firth S, Labow G, Partyka G, Pawson S, Reale O, Schubert SD, Suarez M. 2015. 'MERRA-2: Initial evaluation of the climate'. NASA/TM–2015-104606/Vol. 43. http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/ (accessed 4 November 2016).

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD. 2006. Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *J. Geophys. Res.* **111**: D12106, doi: 10.1029/2005JD006548.

Bromwich DH, Nicolas JP, Monaghan AJ, Lazzara MA, Keller LM, Weidner GA, Wilson AB. 2013. Central West Antarctica among the most rapidly warming regions on Earth. *Nat. Geosci.* **6**: 139–145, doi: 10.1038/ngeo1671.

Christy JR, Norris WB, Spencer RW, Hnilo JJ. 2007. Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J. Geophys. Res.* **112**: D06102, doi: 10.1029/2005JD006881.

Cowtan K, Way RG. 2014. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**: 1935–1944, doi: 10.1002/qj.2297.

Cowtan K, Hausfather Z, Hawkins E, Jacobs P, Mann ME, Miller SK, Steinman BA, Stolpe MB, Way RG. 2015. Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.* **42**: 6526–6534, doi: 10.1002/2015GL064888.

Davy R, Esau I. 2016. Differences in the efficacy of climate forcings explained by variations in atmospheric boundary-layer depth. *Nat. Commun.* **7**: 11690, doi: 10.1038/ncomms11690.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597, doi: 10.1002/qj.828.

Dee DP, Balmaseda M, Balsamo G, Engelen R, Simmons AJ, Thépaut J-N. 2014. Toward a consistent reanalysis of the climate system. *Bull. Am. Meteorol. Soc.* **95**: 1235–1248, doi: 10.1175/BAMS-D-13-00043.1.

Donlon CJ, Martin M, Stark J, Roberts-Jones J, Fiedler E, Wimmer W. 2012. The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens. Environ.* **116**: 140–158, doi: 10.1016/j.rse.2010.10.017.

Fréville H, Brun E, Picard G, Tatarinova N, Arnaud L, Lanconelli C, Reijmer C, van den Broeke M. 2014. Using MODIS land surface temperatures and the Crocus snow model to understand the warm bias of ERA-Interim reanalyses at the surface in Antarctica. *Cryosphere* **8**: 1361–1373, doi: 10.5194/tc-8-1361-2014.

Fyfe JC, Meehl GA, England MH, Mann ME, Santer BD, Flato GM, Hawkins E, Gillett NP, Xie S-P, Kosaka Y, Swart NC. 2016. Making sense of the early-2000s warming slowdown. *Nat. Clim. Change* **6**: 224–228, doi: 10.1038/nclimate2938.

GCOS. 2015. *Status of the Global Observing System for Climate*. GCOS Publication No. 195. WMO: Geneva, Switzerland. http://www.wmo.int/pages/prog/gcos/Publications/GCOS-195_en.pdf (accessed 4 November 2016).

Gemmill W, Katz B, Li X. 2007. 'Daily real-time global sea surface temperature – High resolution analysis at NOAA/NCEP'. NOAA/NWS/NCEP/MMAB Office Note 260. http://polar.ncep.noaa.gov/mmab/papers/tn260/ (accessed 4 November 2016).

Gleason B, Williams C, Menne M, Lawrimore J. 2015. 'Modifications to GHCN-monthly (version 3.3.0) and USHCN (version 2.5.5) processing systems'. GHCN-M Technical report GHCNM-15-01. NOAA NCEI, Asheville, NC. http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20Report%20GHCNM%20No15-01.pdf (accessed 4 November 2016).

Hansen JE, Lebedeff S. 1987. Global trends of measured surface air temperature. *J. Geophys. Res.* **92**: 13345–13372, doi: 10.1029/JD092iD11p13345.

Hansen J, Ruedy R, Sato M, Lo K. 2010. Global surface temperature change. *Rev. Geophys.* **48**: RG4004, doi: 10.1029/2010RG000345.

Hartmann D, Klein Tank AMG, Rusticucci M, Alexander LV, Brönnimann S, Charabi Y, Dentener FJ, Dlugokencky EJ, Easterling DR, Kaplan A, Soden BJ, Thorne PW, Wild M, Zhai PM. 2013. Observations: Atmosphere and surface. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. (eds.): 159–254. Cambridge University Press: Cambridge, UK and New York, NY.

Hersbach H, Peubey C, Simmons A, Berrisford P, Poli P, Dee D. 2015. ERA-20CM: A twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.* **141**: 2350–2375, doi: 10.1002/qj.2528.

Hirahara S, Balmaseda MA, Hersbach, H. 2016. 'Sea surface temperature and sea ice concentration for ERA5', ERA Report Series no. 26, ECMWF: Reading, UK.

Huang B, Banzon VF, Freeman E, Lawrimore J, Liu W, Peterson TC, Smith TM, Thorne PW, Woodruff SD, Zhang H-M. 2015. Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Clim.* **28**: 911–930, doi: 10.1175/JCLI-D-14-00006.1.

IPCC. 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL. (eds.): 996 pp. Cambridge University Press: Cambridge, UK and New York, NY.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. (eds.): 1535 pp. Cambridge University Press: Cambridge, UK and New York, NY.

Ishii M, Shouji A, Sugimoto S, Matsumoto T. 2005. Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the KOBE collection. *Int. J. Climatol.* **25**: 865–879, doi: 10.1002/joc.1169.

Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M, Morice CP. 2012. Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.* **117**: D05127, doi: 10.1029/2011JD017139.

Karl TR, Arguez A, Huang B, Lawrimore JH, McMahon JR, Menne MJ, Peterson TC, Vose RS, Zhang H-M. 2015. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**: 1469–1472, doi: 10.1126/science.aaa5632.

Kennedy JJ, Rayner NA, Smith RO, Saunby M, Parker DE. 2011. Reassessing biases and other uncertainties in sea-surface temperature observations since 1850, part 2: Biases and homogenisation. *J. Geophys. Res.* **116**: D14104, doi: 10.1029/2010JD015220.

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K. 2015. The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.* **93**: 5–48, doi: 10.2151/jmsj.2015-001.

Knutti R, Rogelj J, Sedlacek J, Fischer EM. 2015. A scientific critique of the two-degree climate change target. *Nat. Geosci.* **9**: 13–18, doi: 10.1038/ngeo2595.

Laloyaux P, Balmaseda M, Dee D, Mogensen K, Janssen P. 2016. A coupled data assimilation system for climate reanalysis. *Q. J. R. Meteorol. Soc.* **142**: 65–78, doi: 10.1002/qj.2629.

Lewandowsky S, Risbey J, Oreskes N. 2015. The 'pause' in global warming: Turning a routine fluctuation into a problem for science. *Bull. Am. Meteorol. Soc.* **97**: 723–733, doi: 10.1175/BAMS-D-14-00106.1.

Masson-Delmotte V, Schulz M, Abe-Ouchi A, Beer J, Ganopolski A, González Rouco JF, Jansen E, Lambeck K, Luterbacher J, Naish T, Osborn T, Otto-Bliesner B, Quinn T, Ramesh R, Rojas M, Shao X, Timmermann A. 2013. Information from paleoclimate archives. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. (eds.): 383–464. Cambridge University Press: Cambridge, UK and New York, NY.

Mears CA, Wentz FJ. 2009. Construction of the remote sensing systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders. *J. Atmos. Oceanic Technol.* **26**: 1040–1056, doi: 10.1175/2008JTECHA1176.1.

Morice CP, Kennedy JJ, Rayner NA, Jones PD. 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res.* **117**: D08101, doi: 10.1029/2011JD017187.

Peterson TC, Willett KM, Thorne PC. 2011. Observed changes in surface atmospheric energy over land. *Geophys. Res. Lett.* **38**: L16707, doi: 10.1029/2011GL048442.

Poli P, Hersbach H, Dee D, Berrisford P, Simmons A, Vitart F, Laloyaux P, Tan D, Peubey C, Thépaut J-N, Trémolet Y, Holm E, Bonavita M, Isaksen L, Fisher M. 2016. ERA-20C: An atmospheric reanalysis of the 20th century. *J. Clim.* **29**: 4083–4097, doi: 10.1175/JCLI-D-15-0556.1.

Rhein M, Rintoul SR, Aoki S, Campos E, Chambers D, Feely RA, Gulev S, Johnson GC, Josey SA, Kostianoy A, Mauritzen C, Roemmich D, Talley LD, Wang F. 2013. Observations: ocean. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. (eds.): 255–316. Cambridge University Press: Cambridge, UK and New York, NY.

Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim G-J, Bloom S, Chen J, Collins D, Conaty A, da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redde CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J. 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**: 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Simmons AJ, Poli P. 2015. Arctic warming in ERA-Interim and other analyses. *Q. J. R. Meteorol. Soc.* **141**: 1147–1162, doi: 10.1002/qj.2422.

Simmons AJ, Jones PD, da Costa Bechtold V, Beljaars ACM, Kållberg PW, Saarinen S, Uppala SM, Viterbo P, Wedi N. 2004. Comparison of trends and low-frequency variability in CRU, ERA-40 and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.* **109**: D24115, doi: 10.1029/2004JD005306.

Simmons AJ, Willett KM, Jones PD, Thorne PW, Dee DP. 2010. Low-frequency variations in surface atmospheric humidity, temperature and precipitation: Inferences from reanalyses and monthly gridded observational datasets. *J. Geophys. Res.* **115**: D01110, doi: 10.1029/2009JD012442.

Simmons AJ, Poli P, Dee DP, Berrisford P, Hersbach H, Kobayashi S, Peubey C. 2014. Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q. J. R. Meteorol. Soc.* **140**: 329–353, doi: 10.1002/qj.2317.

UNFCCC. 2015. 'The Paris Agreement reached by Parties to the United Nations framework convention on climate change'. http://unfccc.int/meetings/paris_nov_2015/items/9445.php (accessed 4 November 2016).

Vose RS, Arndt D, Banzon VF, Easterling DR, Gleason B, Huang B, Kearns E, Lawrimore JH, Matthew J, Menne MJ, Peterson TC, Reynolds RW, Smith TM, Williams CN Jr, Wuertz DB. 2012. NOAA's merged land–ocean surface temperature analysis. *Bull. Am. Meteorol. Soc.* **93**: 1677–1685, doi: 10.1175/BAMS-D-11-00241.1.